

AD624144

AMRL-TR-65-102

A REVIEW OF THE EFFECTS OF HIGH AMBIENT TEMPERATURE ON MENTAL PERFORMANCE

JOHN F. WING

CLEARINGHOUSE FOR FEDERAL SCIENTIFIC AND TECHNICAL INFORMATION			
Hardcopy	Microfiche		
2.00	\$0.50	4000	02
ARCHIVE COPY			

SEPTEMBER 1965

Code 1

AEROSPACE MEDICAL RESEARCH LABORATORIES
AEROSPACE MEDICAL DIVISION
AIR FORCE SYSTEMS COMMAND
WRIGHT-PATTERSON AIR FORCE BASE, OHIO

DISC 100
JISA E

AMRL-TR-65-102

**A REVIEW OF THE EFFECTS OF HIGH AMBIENT
TEMPERATURE ON MENTAL PERFORMANCE**

JOHN F. WING

Foreword

This review was prepared in the Environmental Stress Branch, Training Research Division, Behavioral Sciences Laboratory, under Project No. 1710, "Training, Personnel, and Psychological Stress Aspects of Bioastronautics," Task No. 171002, "Performance Effects of Environmental Stress." Dr. Gordon A. Eckstrand served as Project Scientist and Dr. Charles S. Harris as Task Scientist. The review was initiated in early 1964 and completed in January 1965.

This technical report has been reviewed and is approved.

WALTER F. GREETHER, PhD
Technical Director
Behavioral Sciences Laboratory
Aerospace Medical Research Laboratories

Abstract

Fifteen experiments done in various laboratories have assessed the effects of high thermal stress on mental performance. These experiments represent different combinations of exposure time and effective temperature. These studies were reviewed, and the upper thermal limit for unimpaired mental performance was found to vary systematically with exposure duration. Specifically, the lowest test temperatures yielding statistically-reliable decrements in mental performance decline exponentially as exposure durations are increased up to 4 hours. When this temperature-duration curve for mental performance is compared with physiological tolerance curves, it is found to lie well below them at every point in time.

Table of Contents

Section No.	Page No.
I. INTRODUCTION	1
II. REVIEW OF TEMPERATURE EFFECTS ON MENTAL PERFORMANCE	2
Durations of Less than One Hour	3
One-Hour Duration	6
Two-Hour Duration	9
Three-Hour Duration	16
Four-Hour Duration	19
Six-Hour Duration	19
Eight-Hour Duration	21
III. CONCLUSION	23
Determination of the Temperature-Duration Function	23
Specification and Generalization of the Temperature-Duration Function	25
APPENDIX	27
REFERENCES	33

List of Figures

Figure No.		Page No.
1	Average, Combined T-score for Eight Subjects at Selected Points During Experiments in Each Environment .	5
2	Average Number of Correctly Recalled Words on Each Trial under Three Effective Temperatures	9
3	Effects of Environmental Warmth and Time Per Comparison on High Speed Decision Taking	10
4	Average Number of Comparisons Missed Per Hundred in Each Climate .	12
5	Average Number of Seconds Per Terminal as a Function of Effective Temperature	13
6	Mean Number of Problems Completed by Four Subjects as a Function of Effective Temperature	15
7	Average Error Rate in Percent at Each of Five Effective Temperatures	17
8	Average Error Rate on a Coding Test at Each of Five Effective Temperatures	18
9	Comparison of the Tentative, Upper Limit for Unimpaired, Mental Performance with Both the Tolerable and Marginal Physiological Limits	24
10	Summary of the Exposure Durations and Effective Temperatures Used in the Fifteen Experiments Reviewed .	31

List of Tables

Table No.	Page No.
1 Average Performance under Each Effective Temperature Condition as Obtained by Bartlett and Gronow	7
2 Average Performance under Each Effective Temperature Condition as Obtained by Chiles in Experiment One	7
3 Average Performance under Each Effective Temperature Condition as Obtained by Chiles in Experiment Two	8
4 Class-Interval Grouping of Seventeen Test Temperatures Used by Givoni and Rim	16
5 Average Number of Correct Anagram Solutions Under Each Effective Temperature Condition as Obtained by Fine, et al	20
6 Average Grades Made by Matched Groups Instructed under Two Different Temperature Conditions as Obtained by Mayo	22
7 Lowest Test Temperatures Yielding Reliable Decrements	23
8 Reorganization of Givoni and Rim Data on Number of Completed Problems	27
9 Each Subject's Average Performance Scores Ranked Across the Five Temperature Levels for Run One	28
10 Each Subject's Average Performance Scores Ranked Across the Five Temperature Levels for Run Two	29

A Review of the Effects of High Ambient Temperature on Mental Performance

SECTION I Introduction

At the present time there exist several reviews of the effects of elevated temperatures on human performance (refs. 3, 9, and 19). These reviews have proved very useful, but they do not establish a very clear picture of the probable thermal limits for unimpaired performance. One of the main reasons for this is that reviewers have failed to consider duration of exposure as a systematic variable of major importance. Indeed when exposure duration is taken into account, high temperature performance data begin to take on a coherence otherwise missing. This fact led the present author to believe it would be possible to establish tentative, upper limits for unimpaired performance. Although such limits would inevitably be found to err at certain points, they would serve as an important guideline for further research and thereby lead to more precise and reliable estimates. Presentation of tentative, upper performance limits, then, will be the purpose of a series of technical reports. This report on the upper limits of mental performance is the first of the series.

SECTION II

Review of Temperature Effects on Mental Performance

In reviewing the effects of high temperatures on mental performance, repeated reference will be made to figure 10 (see p. 31). Although figure 10 is a summary of the experimental findings reviewed in this report, it also serves as an effective guide to the experimental conditions of each of the 15 studies reviewed here. Figure 10 appears on the final page of the report and has been designed so that it folds out; as a result, it can be readily referred to during the entire reading of the review. The figure shows the test temperatures and exposure times for all 15 experiments. The abscissa for figure 10 is exposure time in minutes. With one or two exceptions, the duration of the experiments ran in exact hourly intervals. It is impractical to plot the test points for several experiments at exactly the same hourly interval, so they have been grouped tightly around the interval instead. This is indicated by the brackets.

The left-hand ordinate for figure 10 is temperature in degrees Fahrenheit as measured on the effective temperature scale. This scale was introduced in 1923 by Houghten and Yaglou (ref. 13). It is an empirically determined index of the degree of warmth experienced by subjects exposed to various combinations of temperature, humidity and air movement.* All but one of the experiments discussed here reported their test temperatures in terms of the effective temperature scale; and in the case of the only exception (Bartlett and Gronow), the data necessary for computing effective temperatures were available in the author's original report (ref. 1, p. 10).

The right-hand ordinate is effective temperature in degrees centigrade. No use has been made of the centigrade scale in the body of this review, however. All the studies discussed here reported their results in the Fahrenheit scale, and it would be more beneficial for readers already familiar with these studies to also have them reviewed in terms of the Fahrenheit scale.

Each vertical line in figure 10 represents a single experiment. The round circles on a given line are the test temperatures. In most experiments subjects were exposed for an identical time period to each of the various test temperatures, so that the line connecting the circles appears vertical on the figure. In one experiment, however, subjects had to be removed from extreme test temperatures after differing durations of exposure; as a result, the line for this study appears as a diagonal line. The letter at the bottom of any given line is the initial of the investigator's last name. Where an experiment is one of several performed by an investigator, it is identified by a number following his initial. (For example, two experiments reported by Mackworth have been labelled M-1 and M-2.) Where a study was reported by two authors, both of their final initials are given (eg, B-G stands for Bartlett and Gronow).

The bold letter at the top of each line keys the reader to the type of mental task used by the investigator. The key to these tasks is printed directly on the figure itself.

Finally, some of the circles representing test temperatures are empty, some are half-filled, and some are solid. Empty circles represent control and experimental temperatures at which no decrement in mental performance occurred. Half-filled circles represent test temperatures at which a decrement was noted by the author, but at which he either applied no statistical test or else used a questionable statistical procedure for establishing the level of significance. Filled or solid circles represent experimental temperatures at which appropriate statistical procedures showed a statistically-reliable decrement in mental performance ($p < .05$ or above).

The experiments represented in figure 10 are reviewed in a left to right order according to their duration of exposure. First to be reviewed is an experiment by Blockley and Lyman (B-L)

*Subsequently, various modifications of the original scale have been introduced, including a correction for radiant heat (see ref. 2).

which deals with brief exposures to very high temperatures. Then, the other studies are considered in separate sections dealing with 1-hour duration, 2-hour duration, 3-hour duration, and so forth. At the end of each section some conclusion is reached as to which studies have provided the best estimate of the temperature threshold for impairment of mental performance at that specific exposure duration. These best estimates for each duration of exposure are then drawn together in the conclusion section, where they are used to construct a temperature-duration curve representing a tentative, upper limit for unimpaired mental performance.

STUDIES OF LESS THAN ONE HOUR

Blockley and Lyman (ref. 4)* exposed eight subjects for brief periods to extremely high temperatures and tested for decrements in mental arithmetic and number checking. The subjects were six Naval Reserve pilots and two amateur, private pilots. The test temperatures were 160°, 200° and 235°F dry bulb with humidity constant at a vapor pressure of 0.8 in. Hg. The authors report that the effective temperature equivalents were 100.5°, 109° and 114°F. The average durations of exposure to these effective temperatures were 24, 36, and 72.5 minutes, respectively, and these varying lengths of exposure account for the irregular shape of the "B-L" line in figure 10. Subjects were removed when physiological indicators showed that their maximum tolerance limits had been reached.

Blockley and Lyman's experimental design did not meet sound principles of counterbalancing, partly because of technical difficulties which invalidated results from three sessions. The design originally called for randomly assigning among 6 subjects the 6 possible orders of exposure to the 3 test temperatures, while the 2 remaining subjects were to be exposed 3 times in succession to the middle temperature of 109° before receiving the other 2 temperatures (each in a different order). The effect of the three incomplete sessions, however, was to postpone that test temperature so that it was repeated out of sequence after all other sessions. These enforced departures from the original design did complicate interpretation of certain results.

Regardless of the order of the three experimental temperatures, each subject always received an initial and a terminal session consisting of a 1-hour test under room temperature conditions which varied between 80° and 90°F dry bulb. Thus, the typical subject had five sessions: the first and last sessions were 80°F control temperatures, and the middle three sessions were his assigned sequence of experimental test temperatures. Subjects were also briefly tested at room temperature before and after each high temperature session. Therefore, control data were obtained not only before and after the experiment proper, but (more importantly) just before and just after each experimental, high temperature session.

In every session (including the two sessions at room temperature) subjects were administered a mental addition task and a number-checking task. Twenty-four pages of addition problems and of number-pairs were used. Pages of mental addition were alternated with pages of number-checking. In both these tasks subjects were allowed 2.75 minutes per test page with a 15-second rest interval between pages. Subjects worked continuously (except for the 15-second rest intervals) from the first minute of exposure until their physiological tolerance limits were reached. The number of pages they had completed, therefore, varied with each session. Subjects received a raw score for each page based on the number of correctly completed problems on that page. The raw score was then converted into a standard score (T-score) based on a separate frequency distribution of the scores made on that page by 105 male students who had been separately administered all the pages used in the experiment. Such a procedure effectively equates the pages

*See also Lyman (ref. 14).

for difficulty. This allows the experimenter to test for temperature-induced shifts in a subject's position in the group.

Blockley and Lyman's results are partially summarized in figure 1. The changes over time in the average T-scores for the eight subjects are shown for each of the three, experimental, high temperature sessions in relation to the two control sessions (the lines labelled 1st 80°F and 2nd 80°F). It also shows the tests made at room temperature just before and just after each high temperature session (points labelled Preexposure and Postexposure). Since there was a different, average exposure period for each temperature, the authors plotted performance as a function of the proportion of exposure time; that is, they plotted performance at quarter intervals of the total exposure period. The ordinate is the combined T-score, i.e., the average of the T-scores for arithmetic and number-checking. The figure shows that at 114°F (ET), this combined T-score drops from a Preexposure T-score of about 96 to a T-score of about 80 at the end of exposure. This drop is highly significant ($p < .01$) when a t test is applied to the difference score. Since Blockley and Lyman report the average duration of exposure under this temperature to be 24 minutes, this would mean that a significant decrement in performance of the combined tasks appears at this temperature prior to an exposure duration of 24 minutes. A decrement actually occurred in both mental addition and number-checking. However, when t tests were applied separately to difference scores for the two tasks, the drop was found to be significant only for the mental addition task ($p < .02$).

At 109°F (ET), the average combined T-score for the Preexposure period is about 94 and this drops by the end of the exposure period to a T-score of 92. This drop is not significant, however. Detailed reasons for this are considered below. At 100.5°F (ET) the T-score drop is from about 94.5 to 82.5, and again a t -test of the difference shows it to be highly significant ($p < .01$). The decrement was significant for both mental addition and number-checking when tested separately. In figure 10, these various results are represented by the filled circles for test temperatures of 114° and 100.5°F (ET), and by the half-filled circle for the 109°F test temperature.

Why did the 109°F temperature fail to yield a statistically-reliable decrement? Lyman (ref. 14, p. 84) remarks that:

"It is of special interest to note . . . that while a statistically significant decrement in performance proficiency was shown on both types of test and in the combination score during the last six minute period in the heat at both the [100.5°F] and the [114°F] temperatures, the decrement at [109°F] was not significant. While this result suggests the possibility that the [109°F] environment might differ in some fundamental way from either the higher or the lower temperatures such a conclusion would be difficult to support, for so far as the author knows, there is no evidence in other stress experiments which shows such a discontinuity as the degree of stress is increased. Actually there is no need to hypothesize such a differential effect to account for the result. In addition to possible uncontrolled biases in the sample that were introduced by the irregularities in procedure at the [109°F] temperature which have already been discussed, it was observed that the pre-exposure score of Subject V was abnormally low as compared to his usual proficiency. The result of this was that he showed improvement at each point in the experiment when his subsequent scores were subtracted from the pre-exposure score. When this subject's performance record was dropped from the data a t of 2.48 ($p < .05$) was obtained for the combination scores of the remaining seven subjects." (The present author has supplied effective temperatures in brackets where the original authors reported dry-bulb readings.)

Nevertheless, the 109°F point in figure 10 is shown as a decrement which was not statistically significant. The reason for doing this is simply one of adhering to strict criteria in reporting sig-

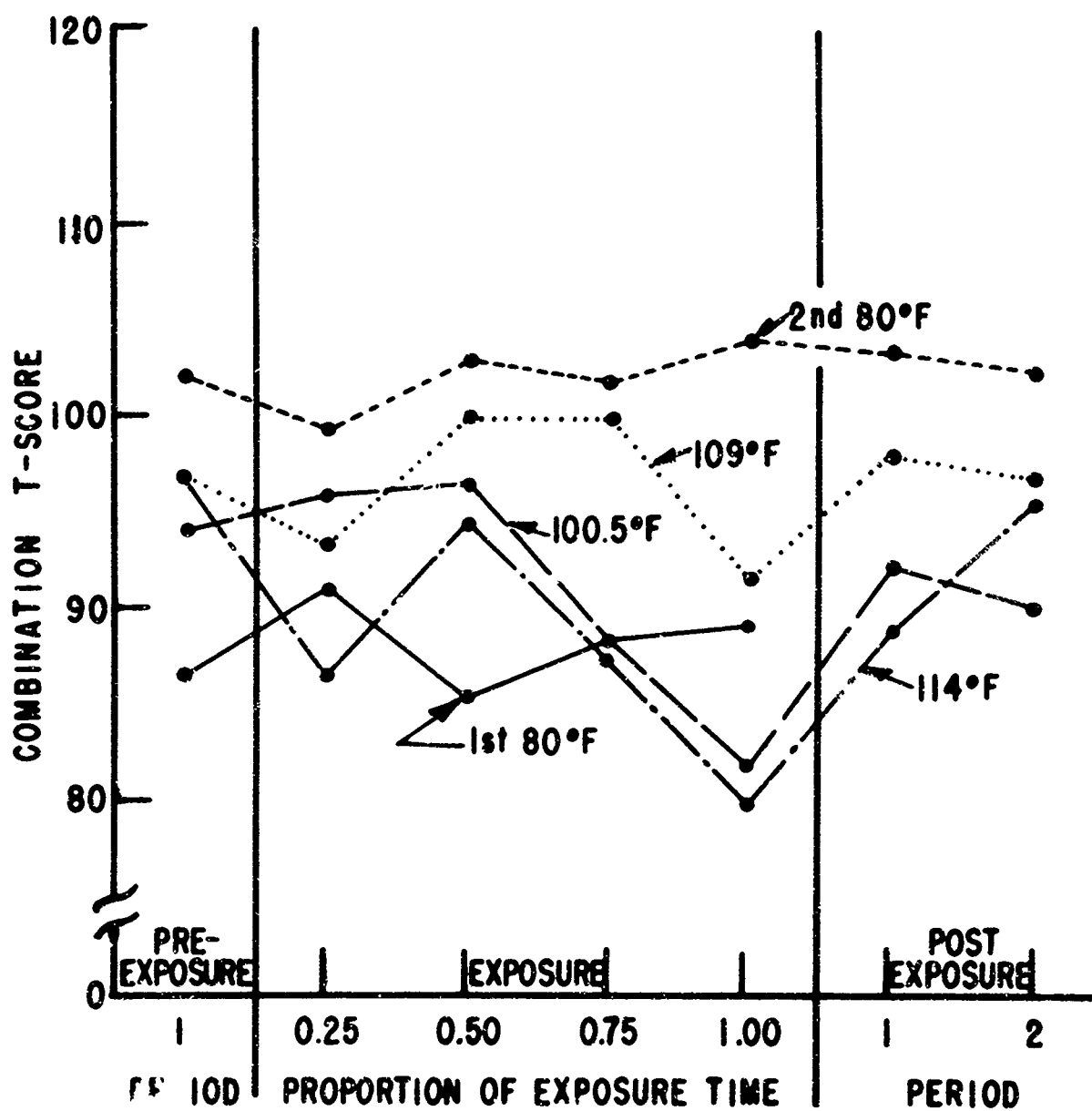


FIGURE 1

Average, Combined T-Score for Eight Subjects at Selected Points During Experiments in Each Environment (Adapted from Blockley and Lyman [ref. 4] and from Lyman [ref. 14, p. 82])

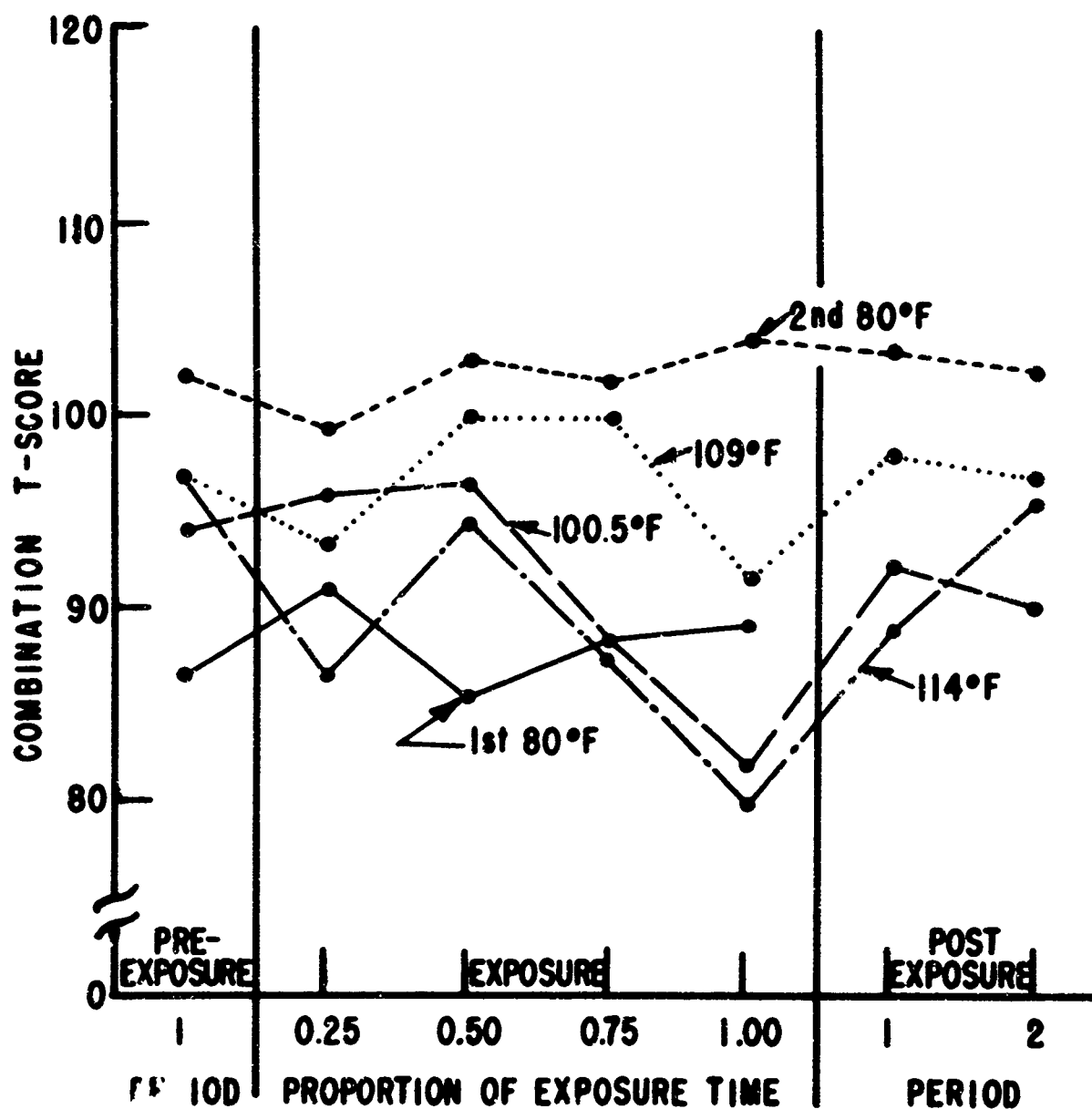


FIGURE 1

Average, Combined T-Score for Eight Subjects at Selected Points During Experiments in Each Environment (Adapted from Blockley and Lyman [ref. 4] and from Lyman [ref. 14, p. 82])

nificant findings. Throughout this report we shall always note those instances in which significance would have been obtained if atypical data were omitted and those instances in which results approached but did not quite reach a significance level of .05. However, such instances will not be considered significant and will not in themselves serve to define the upper limits of unimpaired performance.

ONE-HOUR STUDIES

As already noted, Blockley and Lyman found a significant decrement in mental arithmetic and number-checking performance during a 1-hour exposure (72.5 min. exposure) to 100.5°F effective temperature. Four additional studies have been reported in which subjects were under thermal stress for approximately 1 hour. These four studies will be used to establish for the 1-hour exposure period the probable upper limit for unimpaired mental performance.

Bartlett and Gronow (ref. 1) employed an unpaced mental task in which sixteen subjects had to estimate the collision courses of a number of colored airplane silhouettes. Information as to speed and course were provided them, and they were to imagine the silhouettes as moving in a specified manner (vertically, horizontally or diagonally) from square to square across a grid or matrix which they had before them. They predicted how many crashes would occur, which planes, if any, would collide and on which squares they would collide. Time to reach a decision was also measured. These four performance measures were obtained under 72.5°, 82°, and 91.5°F effective temperatures (as well as under a room temperature condition which ranged between 60° - 70°F dry bulb). Four groups of four subjects each underwent the conditions in a Latin square design to counteract learning and fatigue effects. The task was always given during the final half hour of the 1-hour exposure periods. Analysis of the data showed no significant effects of temperature on any of the four performance measures. The analysis of each measure (an analysis of variance) is not described in sufficient detail to assess the adequacy of the statistical procedures employed. However, Bartlett and Gronow do report the mean scores for each temperature condition, and visual inspection shows no systematic relationship between temperature and error scores for the first three measures. This is evident from table 1. A slight trend toward shorter decision times at higher temperatures is present, but this was apparently not significant. For an unpaced, mental task of this type, then, there appears to be no decrement for one-hour exposures up to an effective temperature of 91.5°F.

Chiles (refs. 6, 7, and 8) used a paced mental task over a comparable range of test temperatures and also found no significant decrement. This suggests that even when a mental task requires speed, no decrement is likely to occur with a 1-hour exposure to this range of temperatures. The task employed by Chiles was a modification of Mackworth's Complex Mental Task (Mackworth, ref. 17). A series of 20 cards located at irregular intervals on a moving loop had to be visually compared to each of 10 stationary cards with respect to the number of differences in the symbols appearing on each. In the first experiment (C-1 of figure 10), Chiles tested eleven students in the final half-hour of an hour-long exposure to 76°, 81°, 86°, and 91°F effective temperature. Four groups of four subjects each underwent the four temperature conditions in different orders according to a Latin square design, so that the effects of learning were equated across temperature conditions. Two measures of performance were taken: frequency of errors and frequency of omissions. Neither errors nor omissions showed any marked, systematic trend as a function of effective temperature. This is shown in table 2. However, frequency of omissions did show a steady and marked decline as a function of experimental sessions, indicating to Chiles the importance of employing a counterbalanced design with this type of task. Omissions actually

TABLE 1
Average Performance Under Each Effective Temperature Condition
 Cited from Bartlett and Gronow (ref. 1, p. 11)

	Decision Time (Secs)	Percent Plane Error Score	Percent Site Error Score	Percent Crash Error Score	Percent Total Error Score
Room					
Temperature	254.9	11.7	10.1	17.3	13.0
71°	258.2	11.5	11.1	18.7	17.1
81°	251.2	10.7	9.4	15.9	12.0
91.5°	247.6	11.5	10.3	17.5	13.1

TABLE 2
Average Performance Under Each Effective Temperature Condition
 Cited from Chiles (ref. 8, p. 92)

	Effective Temperature			
	76°	81°	86°	91°
Omissions	37.7	34.6	37.1	39.8
Errors	10.0	10.4	11.3	10.8
% Errors	6.6	6.8	7.5	7.3

dropped over 30 percent. When the omissions data were subjected to an analysis of variance, the analysis showed a significant effect of sessions ($p < .025$) but no significant effect due to temperature. This first experiment of Chiles, then, confirms Bartlett and Gronow's study and places the performance impairment threshold for the 1-hour duration somewhere above 91°F effective temperature.

A second experiment by Chiles (C-2 in figure 10) should, in principle, have provided the data necessary to estimate the 1-hour impairment threshold; for in this experiment, Chiles not only used test temperatures of 75°, 81°, 86°, and 91°F (ET), but he also extended the range of test temperatures upward by testing five subjects at 109°F effective temperature. However, of the five subjects he tested at 109°F (ET), only three were able to complete the session. A marked increase in errors resulted, but no test of significance could be made because of the small number of subjects. This left ten subjects who were tested on the same symbol-comparison task at each of four effective temperatures: 75°, 81°, 86°, and 91°F. However, due to difficulties in scheduling, only two orders of treatments could be run. Five subjects received the temperatures in the order: 75°-86°-81°-91° and the other five in the order 86°-75°-81°-91°. Chiles obtained quite small differences among the various temperature conditions for both errors and omissions, and there was no noticeable trend evident in the data. This is apparent from table 3. Furthermore, he found no significant effect of temperature when the omissions data were subjected to an analy-

sis of variance. Since practice effects were not completely counterbalanced, however, it is doubtful whether Chiles' design provides a completely adequate test of temperature effects. The 91°F temperature condition, for example, appeared in the final or fourth session for both groups of subjects. Since in Chiles' first experiment the number of omissions had dropped over 30 percent by the fourth session, it is not unlikely that a similar drop in omissions may have occurred in this experiment. As a result, any moderate but significant increase in omissions produced by the 91°F temperature could have been overshadowed by the large reduction in omissions due to practice. Chiles reports (personal communication) that six, preliminary, practice trials were administered in this experiment so that the effects of practice over sessions may have been markedly reduced relative to his first study. Nevertheless, the only temperature conditions which can be properly

TABLE 3
Average Performance Under Each Effective Temperature Condition
Cited from Chiles (ref. 8, p. 94)

	Effective Temperature				
	75°F	81°F	86°F	91°F	109°F
Omissions					
Mean	29.6	29.8	30.6	27.1	22.0
SD	7.7	8.2	8.3	9.3	
Errors					
Mean	6.6	5.5	6.9	6.7	16.3
SD	0.82	0.79	0.92	1.4	

evaluated in this design are those of 75° and 86°F (ET), since these were counterbalanced over the first two sessions. While both omissions and errors do increase with this rise in temperature (see table 3), the increase is exceptionally small and presumably not statistically significant. This is not surprising, of course, since his earlier study indicated that the one-hour threshold for impairment lies above 91°F (ET).

The final one-hour study to be considered was performed by the authors, Wing and Touchstone (ref. 27), who used a paced, short-term memory task. Fifteen subjects were tested in three groups of five subjects each.* The three groups were exposed over three successive days to 72°, 90°, and 95°F effective temperature in counterbalanced order, so that practice effects were equally distributed over the temperature conditions. Three equated sets of word-lists were administered in the same order to all three groups, one set each day, so that all the sets were represented under all the temperature conditions on each day. Every set consisted of six-word lists, each of which was presented orally five times in succession, with opportunity for immediate, free recall of the list after each presentation. Subjects began the test 10 minutes after entering the heat chamber and worked continuously for 50 minutes. An analysis of variance of the number of words correctly recalled showed that temperature produced a highly significant effect ($p < .01$), this effect being due to a systematic decline in the number of words correctly recalled with each increase in temperature. This reduction in number of words correctly recalled at the higher temperatures was found on every trial. Figure 2 shows the acquisition curves for the subjects under each tempera-

*Eighteen subjects were actually tested but the scores of three subjects were discarded to equalize the sample sizes of the three groups in order to employ an analysis of variance.

ture condition. On all five trials of immediate recall, the average number of correct words was highest for 72°, next highest for 90° and lowest for 95°F (ET). This suggests that subjects would have reached criterion in fewer trials at the lower temperatures. (Unfortunately the author did not continue testing subjects under each temperature until perfect performance had been reached.)

Application of "t" tests to the mean scores of the temperature conditions showed that the only statistically-significant difference ($p < .05$) was between the 72° and 95°F temperatures. The 95° test temperature on line W-T in figure 10 is therefore shown as a filled circle. The 90° decrement was not statistically-reliable.

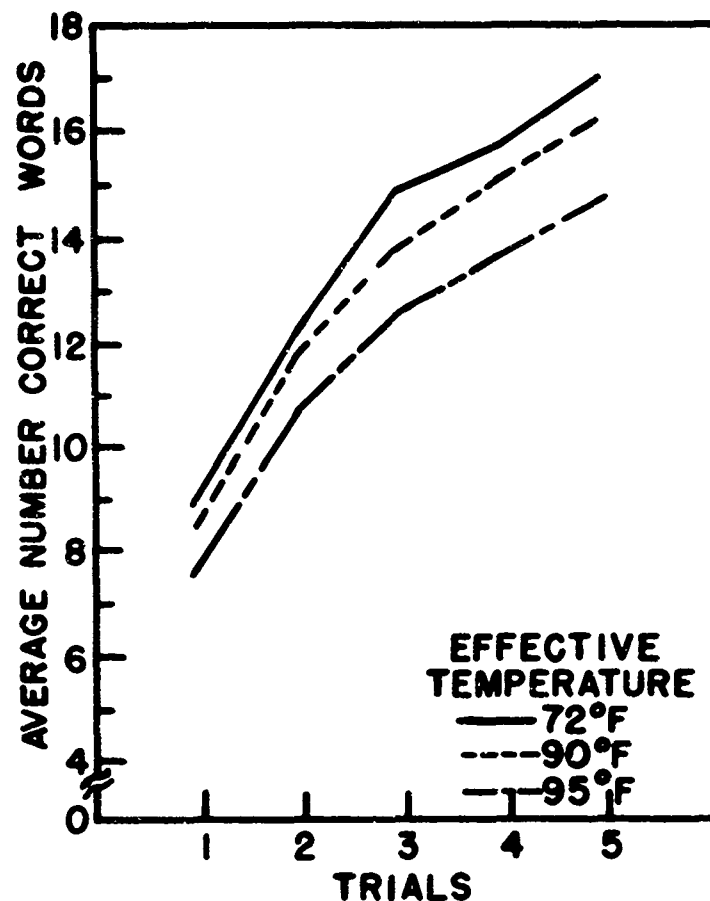


FIGURE 2

Average Number of Correctly Recalled Words on Each Trial Under Three Effective Temperatures (Adapted from Wing and Touchstone [ref. 27])

The results of Wing's experiment, then, suggest that the 1-hour threshold lies between 90° and 95°F (ET). Furthermore, the fact that no statistically-reliable decrements were obtained by Bartlett and Gronow at 91.5°F or by Chiles (in C-1) at 91°F suggests that the upper limit lies somewhere between 91.5° and 95°F (ET).

TWO-HOUR STUDIES

Four studies of mental performance under 2-hour exposure have been performed. Two of the studies used the same symbol-comparison task that Chiles employed above. These symbol-

comparison studies were performed by Pepler (refs. 21 and 22) and are labelled P-1 and P-2 in figure 10. In both of them Pepler used naturally-acclimitized subjects who had been in the tropics for 6 months or longer. Subjects were exposed to 76°, 81°, 86°, and 91°F effective temperature. Pepler's experimental designs for both experiments were well constructed for testing the main effects of temperature conditions, since practice was spread equally over all temperatures by the use of a Latin square design. In his first experiment (P-1), subjects were tested the last hour and a half of the 2-hour period. He required 24 subjects to perform the symbol-comparison task at each of three speeds: slow, medium, and fast. Figure 3 shows Pepler's results in terms of the average number of omissions for each speed under all temperatures. Only for the slow speed did Pepler find a significant increase ($p < .05$) in the number of omissions, and the increase he found occurred at 76°, 86°, and 91°F (ET) over the number of omissions made at an apparently optimal temperature of 81°F (ET). But in testing for significance, Pepler subjected the data for each speed to a separate analysis of variance. Properly, an overall analysis of variance should have been employed so that the three levels of speed would have been tested as a single source

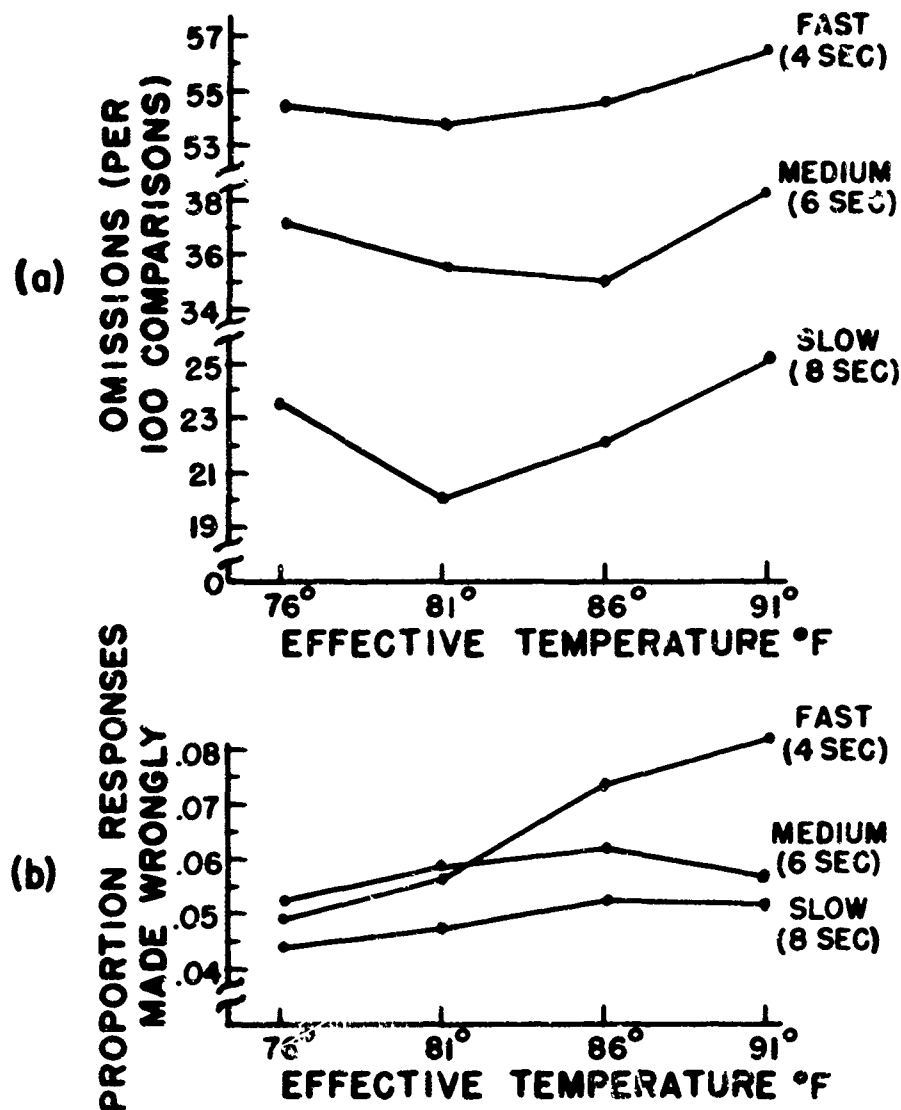


FIGURE 3
Effects of Environmental Warmth and Time per Comparison on High Speed Decision Taking (Adapted from Pepler [ref. 23, p. 85])

of variance and also in interaction with the other sources of variance. Pepler's conclusions are not completely justified without the support of such an overall analysis, and so we must conclude that no statistically-reliable differences among the temperature treatments have been demonstrated.

Pepler also examined the proportion of comparisons in which errors were made (see figure 3b), and reports a significant increase ($p < .05$) as a function of temperature but only for the fast speed. Again this result was based on application of "t" tests without first applying an overall F test. Pepler's results, then, are suggestive at best and are restricted to only certain experimental conditions. We have indicated this by showing the test temperatures of 76°, 86° and 91°F as half-filled circles in figure 10.

Pepler's second experiment (P-2) used 16 men who were naturally-acclimatized to the tropics for 6 months. Subjects were tested almost continuously over the 2-hour period on the symbol-comparison task under the same four temperatures used in experiment P-1. Once again a Latin square design was used to control for practice effects, but, in addition, the design possessed two levels of incentive and two levels of speed stress. These latter treatments were always counter-balanced across subjects within each session. This is an elaborate design requiring an expanded analysis of variance design to test for the many possible interactions between treatments. However, Pepler did not employ such an analysis, so it is impossible to assess which interactions were significant. Nevertheless, from Pepler's results we see that interactions were present, and that one of the effects of these interactions was the obliteration of any simple monotonic increase in omissions with temperature. Figure 4 shows the functions obtained by Pepler. Pepler's analysis (which did not include a proper assessment of interactions) again showed significant temperature effects ($p < .05$ or better) on rate of omissions for the slow speed (but not for the fast speed). It also showed a significant ($p < .05$ or better) temperature effect for the low incentive condition (but not for the high incentive condition). As in the first study, he also reports a significant effect of temperature ($p < .05$ or better) on the proportion of comparisons in which errors were made, but again only for the fast speed. Although some significant effects of temperature may be present in this study, the statistical procedures do not justify firm conclusions. Nor does the study help even indirectly to determine the probable threshold for unimpaired performance, for any significant temperature effects he may have obtained are unique to particular levels of speed and incentive. They cannot be generalized.

Another 2-hour experiment (labelled "CC" in figure 10) is by Carpenter (ref. 5). It provides more adequate evidence than do Pepler's studies as to the probable performance threshold for a 2-hour exposure to high temperatures. Carpenter used a problem-solving, performance test called the Resistance Box Task. Subjects had to trace out a simple circuit (containing only resistances) by using a circuit diagram and a resistance meter. Sixteen military subjects performed the task during the final 45-minutes of a 2-hour exposure to each of four effective temperatures: 79°, 83°, 92°, and 97°F. Subjects were divided into four groups each containing four subjects. Each group took the temperatures in one of four different orders according to a Latin square design, so that effects of learning were equated for all the temperatures. The basic measure of performance was average time to solution. Carpenter's results showed a systematic increase in solution time with a rise in effective temperature. This is shown in figure 5. Analysis of variance showed the effects of temperature to be significant ($p < .01$). Carpenter also estimated the lowest temperature at which a statistically-reliable impairment in performance would have occurred in this experiment. He did this by extrapolating from a curve fitted to his experimental data points (see figure 5). The value he obtained was 89.2°F (ET).

Of these three studies of two-hour exposure to high temperatures, the Carpenter study so far provides the clearest guidance in setting the probable threshold for impairment on mental tasks. Carpenter's design, statistical procedures, and results are all straightforward. The present author

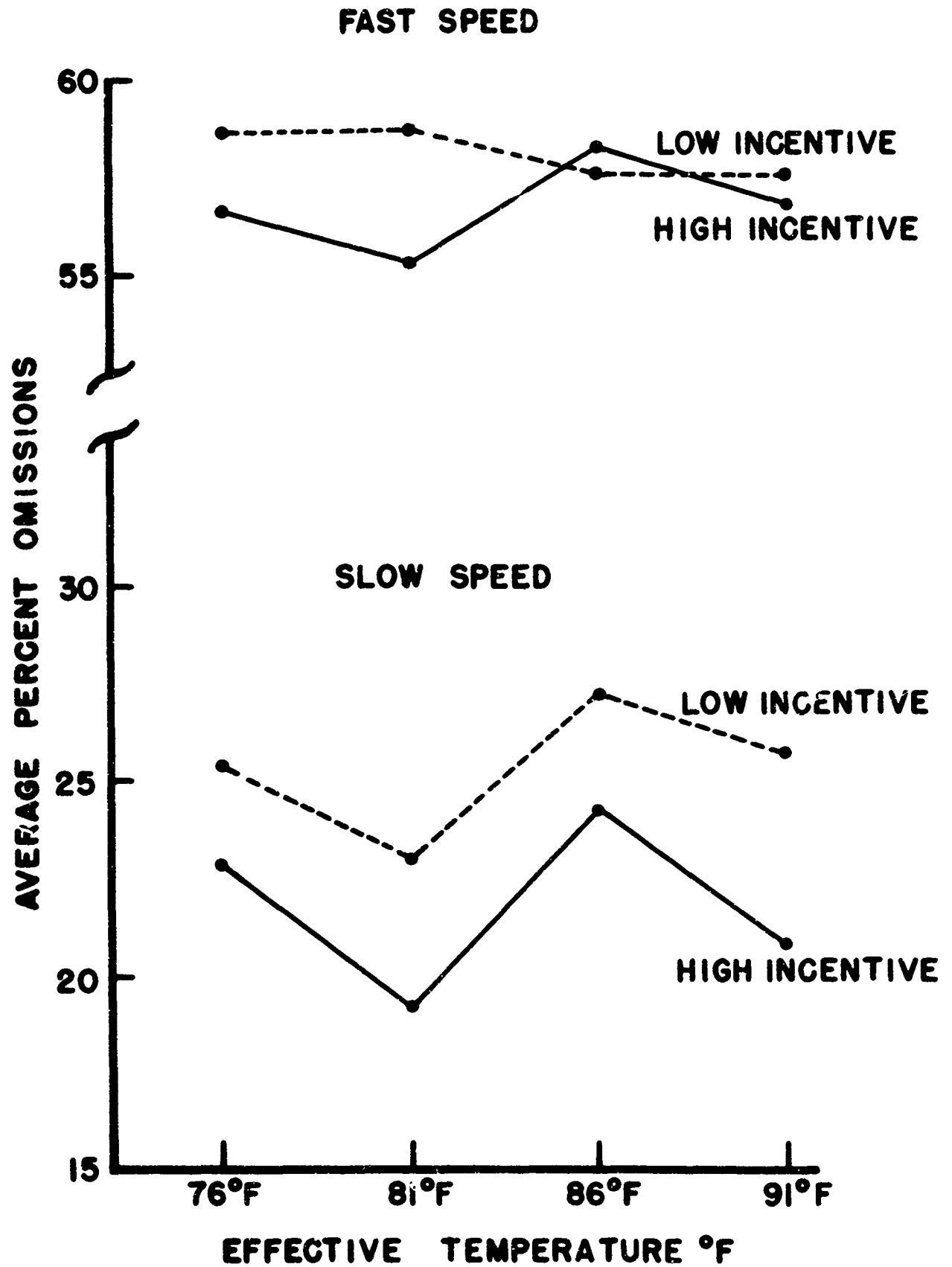


FIGURE 4
Average Number of Comparisons Missed per Hundred in Each Climate
(Adapted from Pepler [ref. 22])

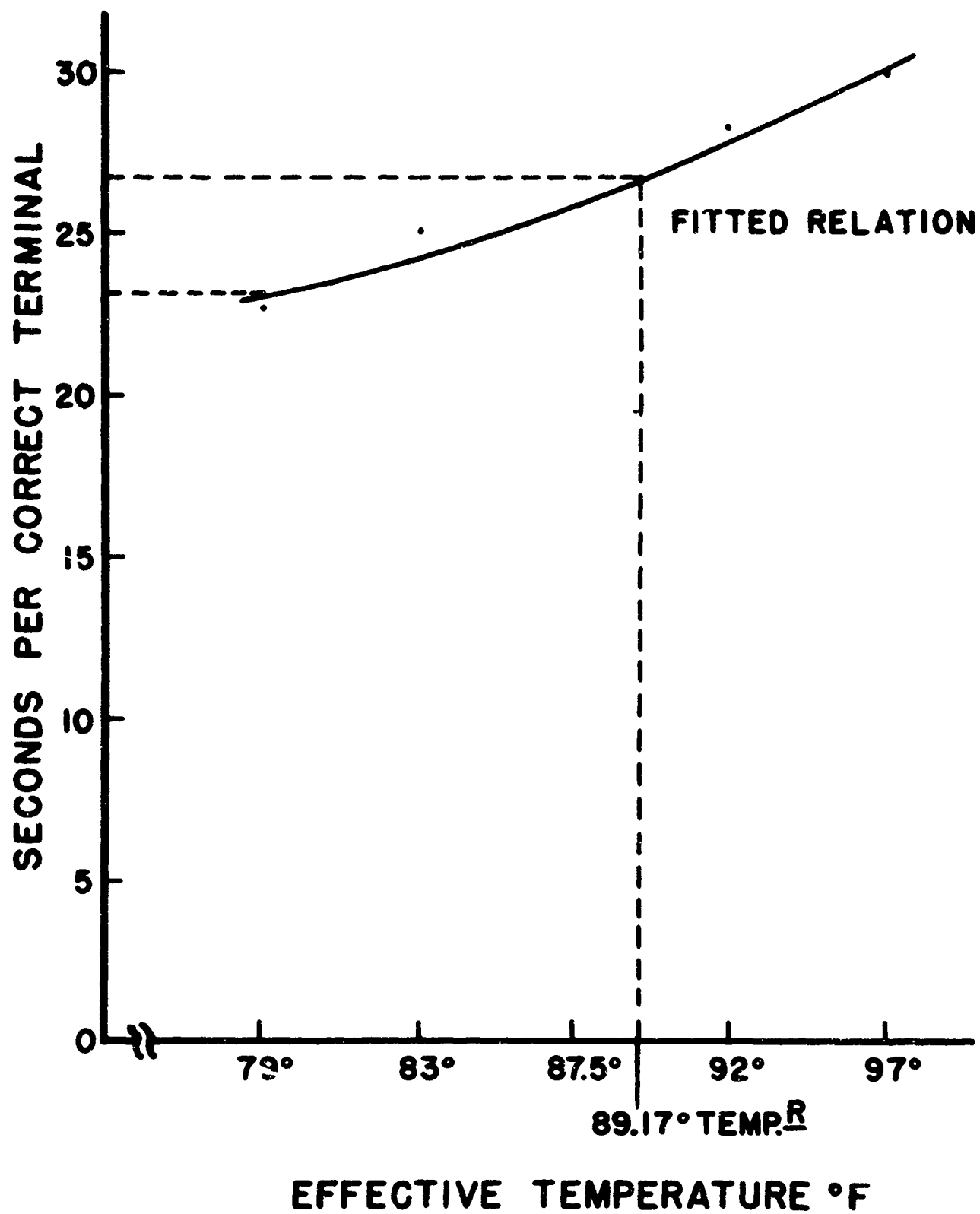


FIGURE 5
Average Number of Seconds per Terminal as a Function of Effective
Temperature (Adapted from Carpenter [ref. 5, p. 25])

is inclined, therefore, to set the tentative upper limit for a two-hour exposure at about 89°F (ET). Pepler's results, of course, would place the threshold lower -- even as low as 86°F (ET) for certain conditions of incentive and speed stress. Under certain conditions with certain tasks, impairment may occur at this low a temperature. However, as already noted, Pepler's statistical procedures make it difficult to accept such a conclusion (even for these special conditions) without further supporting evidence.

Carpenter's figure of 89°F (ET) emerges as a reasonable value for the 2-hour threshold not only on the basis of his own data, but also on the basis of the results of Pepler's first study. In this study (P-1), the temperature which yielded the largest, average number of omissions under all three speeds was 91°F (ET). Whatever the shape of the performance function below this value of 91°F (ET) (that is, regardless of where the optimal temperature point may be at any given speed) at all three speeds the greatest deterioration was at 91°F (ET). Pepler's results too suggest that a reasonable estimate for the 2-hour threshold is some value approaching 91°F.

The final study, a study by Givoni and Rim (ref. 11), investigated four subjects' performance on a mental multiplication task under seventeen test temperatures which ranged between 70.2° and 90.1°F effective temperature. The seventeen temperatures within this range were so presented at various points during the 17 test days as to roughly balance practice effects across temperature conditions. The test days were spread over a period of about 3 weeks. All tests were given in the afternoon. Subjects performed twice during each 2-hour session: run 1 occurred in approximately the final 30 minutes of the first hour and run 2 occurred in the final 30 minutes of the second hour. In each of these 30-minute runs, S's attempted to do as many mental multiplications as possible and were paid on that basis. Each problem consisted of multiplying one 5 digit number by another 5 digit number. The number of problems completed and the number of errors made were recorded under each temperature condition. Givoni and Rim report their main performance results in the form of three figures. One figure shows the appreciable effects of practice on both percent errors and number of problems done over the 17 experimental test days. Two other figures show no relationship between these performance measures and either sweat rate or a thermal sensation index. As a result, the authors conclude that there was no impairment of performance.

However, in their analysis Givoni and Rim simply presented scattergrams. They did not average out practice effects, nor did they make an analysis of performance in terms of the effective temperature scale. Since they published their raw data in the same article, the present author was able to reanalyze the performance data in terms of the effective temperature scale while also using some measure of control over practice effects. The full details of this analysis are presented in Appendix I. In brief, the 17 test temperatures were grouped into five class-intervals of temperature as shown in table 4, and then mean performance was plotted as a function of the effective temperatures representing the midpoints of these class-intervals. Figure 6 shows the average number of completed multiplication problems plotted as a function of the midpoints of these five temperature ranges. These midpoints are 70°, 75°, 80°, 85° and 90°F (ET). (It is these midpoints which are plotted as the test temperatures for line G-R in figure 10). The greatest number of problems was performed in the range of temperatures represented by 75°F; the fewest number of problems were performed in the range of temperatures represented by 90°F. The Friedman two-way analysis of variance, a nonparametric test applied to the ranks of individual subject's scores across treatments, showed a significant overall effect of the temperature treatments ($p < .01$).^{*} Another nonparametric test, the Walsh test, was used to determine which

^{*}A similar analysis of percent errors revealed no significant effects of temperature. The error rate was quite variable and did not show any systematic trend with increasing temperatures.

temperatures produced the obtained significance. Unfortunately, with a sample of only four subjects ($n=4$), the highest level at which the null hypothesis may be rejected using a one-tailed Walsh test is $p<.062$, which is below the .05 level adopted in this paper as the criterion for statistical reliability. We shall, therefore, refer to results significant at $p<.062$ as "almost significant." The test showed that, in both runs, the number of problems completed under 85° and 90°F temperatures were almost significantly fewer ($p<.062$) than the number completed under the 75° temperature. Indeed, in the second run the number of completions under the 90° F temperature was almost significantly less ($p<.062$) than for every one of the other temperatures. While the midpoint of this temperature range is 90°F, the actual mean of the test temperatures in this range is 89.2°F (ET). This value, coincidentally, is the same value which Carpenter esti-

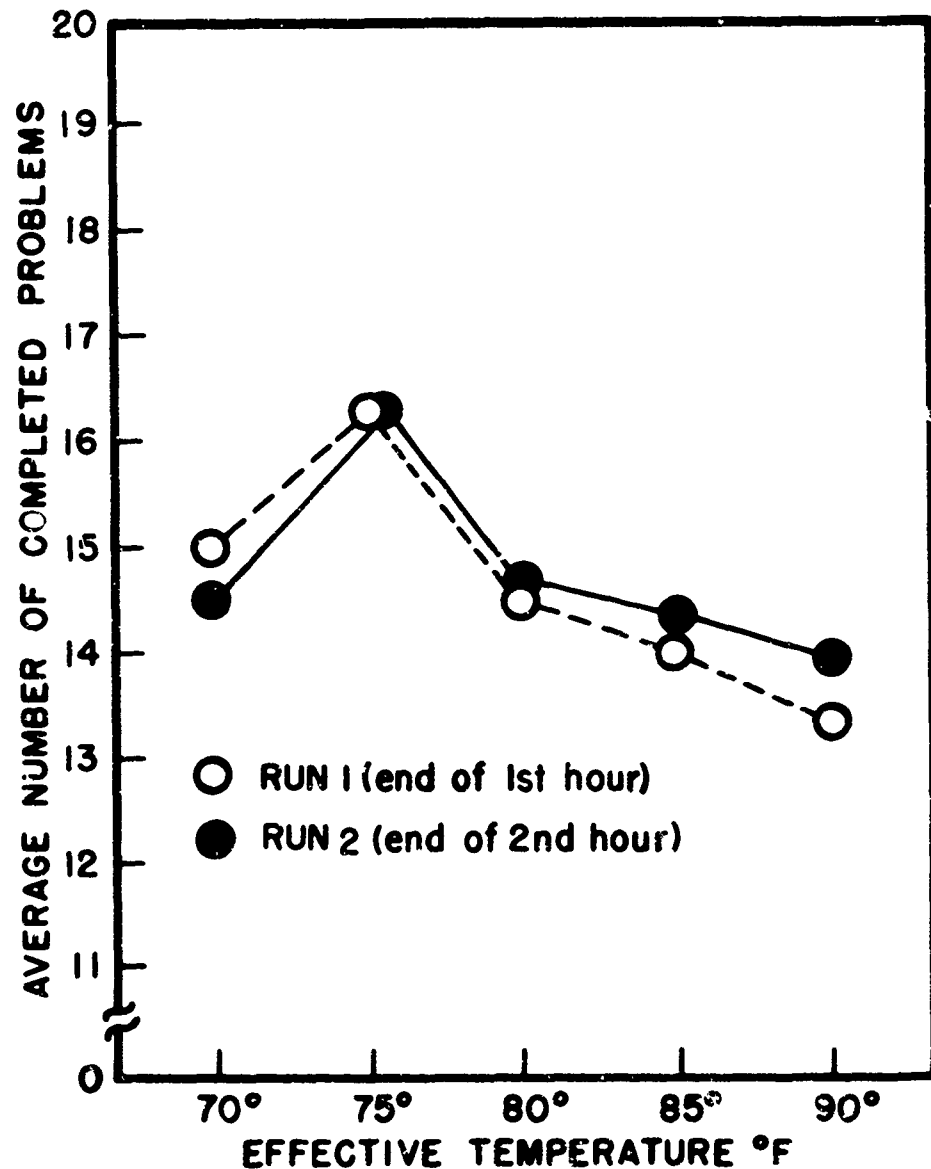


FIGURE 6

Mean Number of Problems Completed by Four Subjects as a Function of Effective Temperature

TABLE 4

*Class-Interval Grouping of Seventeen Test
Temperatures of Givoni and Rim*

Midpoint	Mean	Class Intervals	Effective Temperatures used by Givoni and Rim	Test Days
70.0	70.2	67.5 - 72.4	70.2	8
75.0	74.0	72.5 - 77.4	73.5, 74.5	14, 11
80.0	79.5	77.5 - 82.4	77.9, 79.0, 80.0, 80.9	4, 7, 9, 10
85.0	85.6	82.5 - 87.4	84.9, 85.2, 85.9, 86.4	3, 6, 11, 12
90.0	89.2	87.5 - 92.4	88.0, 88.5, 89.0, 89.5 89.8, 90.1	1, 2, 5, 15, 16, 17

mated by extrapolation to be the lowest temperature at which a statistically reliable impairment occurred in his experiment. In the absence of more definitive data, this value of 89°F (ET) will be taken as the threshold for mental impairment for the 2-hour exposure period. Of course, the basis for selecting 89°F (ET) must be the Carpenter study since the results of Givoni and Rim only approach significance. Nevertheless, their results do provide qualitative support.

THREE-HOUR STUDIES

Three experiments of 3-hour exposure duration have been performed. Two of these are experiments by Mackworth (ref. 17). The first of these (M-1) is a study of the performance of eleven, highly-practiced, telegraphers on a Wireless Telegraphy Test under 79°, 83°, 87.5°, 92°, and 97°F (ET). All subjects had been artificially acclimatized in daily sessions for 2 to 3 months prior to the experiment during which time they also had intensive daily practice at telegraphy. Details of the experimental design are not provided. Although effects of learning would be minimal or nonexistent for such highly-practiced subjects, it would be desirable, of course, to know if other effects of repeated sessions had been held constant either by means of counterbalancing or else by randomly assigning the order of treatments. Subjects received nine messages in each 3-hour session. Each message consisted of 250 groups of five letters and numbers mixed at random. Mackworth tallied the incidence of faulty messages (i.e., any messages with all five symbols wrong or missing), and he found that the average incidence of faulty messages increased as a function of the test temperatures (see figure 7).

A fine-grained analysis was also performed in which tallies were made of any individual letters or numbers which were missing or wrong. The average number of such errors per subject per hour was found to increase for the five temperatures, as follows: 12.0, 11.5, 15.3, 17.3 and 94.7. Mackworth found that "the slight difference between the average error score at the effective temperature of 79°F and that at 83°F could have arisen from chance variations in the experiment. But the increased number of mistakes between the scores at the effective temperatures of 79° and 87.5°F were statistically definite, as also was the rise between the error score made at 83°F and that at 87.5°F" (ref. 17, p. 136). Thus, the lowest temperature at which a reliable decrement was obtained was 87.5°F (ET). This appears as the lowest filled circle on line M-1 in figure 10. This first study by Mackworth, then, clearly suggests that the 3-hour threshold lies at or below 87.5°F (ET).

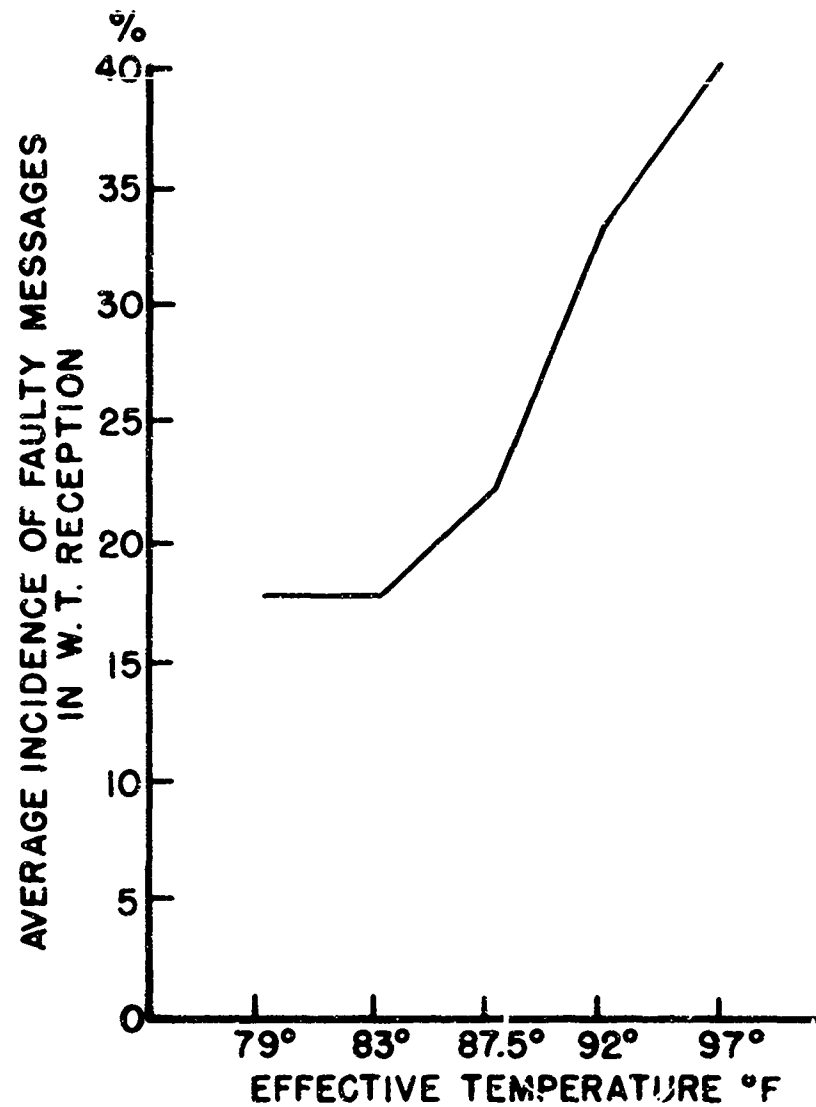


FIGURE 7

Average Error Rate in Percent at Each of Five Effective Temperatures
(Adapted from Mackworth (ref. 17, p. 310))

Pepler (ref. 20) replicated the Mackworth study using twelve, naturally-acclimatized subjects who had been in the tropics 6 months or longer. They were experienced telegraphers. He also gave them further training over a period of 4½ weeks prior to the experiment. He used effective temperatures of 71°, 76°, 81°, 86°, 91°, and 96°F (see P-3 on figure 10). Two groups of six subjects each worked in each of the six temperatures in both morning and afternoon sessions. Subjects performed over the entire 3-hour period. Results showed a significant rise in mean error scores above 86°F for the morning sessions, i.e., 91° and 96°F had significantly higher error incidence than did 86°F. In the afternoon sessions, 91°F had significantly higher error scores than either 81° or 76°F. Pepler's study confirms Mackworth's finding that a decrement in telegraphy performance occurs under high temperatures, but suggests that the upper limit may be somewhat higher for men who are naturally-acclimatized to the tropics.

In Mackworth's second experiment (ref. 17, pp. 141-143) he used a Coding Test (see M-2 on figure 10). This presumably involved more thinking or problem-solving ability than the telegraphy test. The Coding Test consisted of a form-board and small, flat squares. These had to be arranged

on the board according to a code. Twelve subjects were given daily, 3-hour acclimitization sessions for 2 to 3 weeks. During these sessions they also practiced the coding task for 1 hour. In the experiment proper, groups of three subjects each were tested twice under each of the five temperatures, the order of the temperatures being independently randomized each time for each group (rather than counterbalanced). They performed throughout the 3-hour periods. Figure 8 shows the average errors per 100 form-boards at each test temperature. Error rate increases systematically with effective temperature. Mackworth did not perform an analysis of variance or overall test of significance. However, the sizeable and systematic increase in errors which he reports clearly indicates an overall effect due to high temperatures. We will accept, then, the

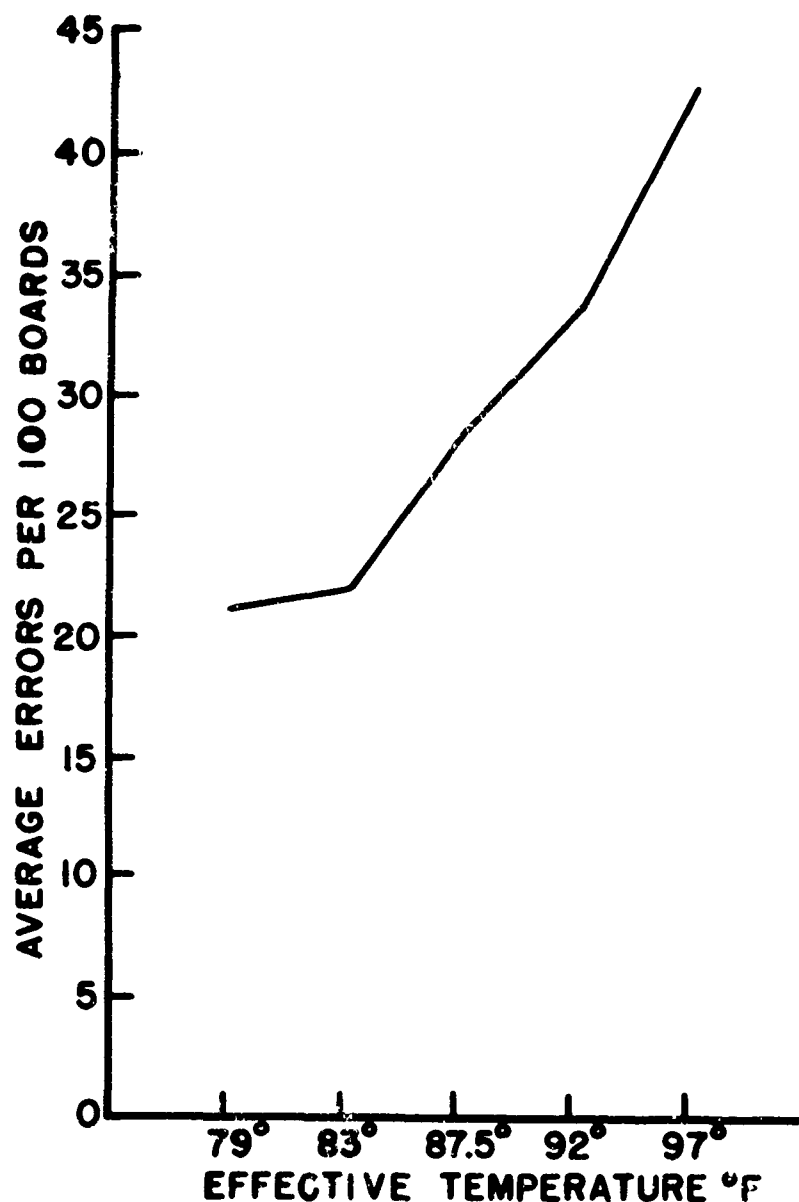


FIGURE 8

Average Error Rate on a Coding Test at Each of Five Effective Temperatures (Adapted from Mackworth [ref. 17, p. 317])

critical ratio tests which he made between temperature conditions. He found that the rise in errors from 79° to 82°F was not significant, but the increases in errors from 79° to 87.5 F and from 83° to 87.5°F were both significant. These findings duplicate exactly the results of the fine-grained analysis of wireless telegraphy errors from his first experiment. This confirmation suggests that the tentative, upper limit for the 3-hour duration lies at 87.5 F (ET) or below.

FOUR-HOUR STUDIES

Viteles and Smith (ref. 25) performed the only study involving a 4-hour exposure (see V-S on figure 10). Their study was performed for the American Society of Heating and Ventilating Engineers as part of a U.S. Navy program, and Viteles and Smith tested those 6 subjects (out of a group of 40 subjects) who seemed to be most homogeneous in regard to the physical and psychological requirements of the U.S. Navy. The subjects were given training on seven tasks during each of four, 2-hour practice sessions. These four sessions were given over four consecutive days and represented four effective temperature conditions: 73°, 80°, 87° and 94 F. (These preliminary sessions, of course, also acted as a brief period of artificial-acclimatization.) The main experiment then tested subjects on all seven tasks under these four temperatures in separate four-hour sessions. It required 42 sessions in all and was scheduled 6 days a week for 7 weeks. One of the seven tests given was definitely mental in nature: the mental multiplication test. This required multiplication of a three-digit number by a two-digit number. Subjects were scored for the number of correct digits in the answer. Another task might also be considered mental: the number checking task which consisted of inspecting pairs of numbers and checking only those pairs which were identical. One-half hour of any given session was devoted to each of these tasks, the ordinal position of each being counterbalanced along with that of the other tasks across the entire experiment.

Viteles and Smith found that none of the subjects could complete the 94°F (ET) condition on first exposure and only four could complete it on the second exposure. They showed marked deterioration in performance, but due to incompleteness in the data it was not analyzed statistically. In analyzing the data from the other three temperature conditions, Viteles and Smith found that for all tests the lowest total output occurred at 87°F (ET). This reduction in output was statistically significant for both the number checking ($p < .05$) and the mental multiplication task ($p < .05$). They did not obtain any significant increase in errors as a function of high temperatures. However, a significant increase ($p < .05$) in variability of scores on the number checking test occurred at the 87°F (ET) level. These findings of Viteles and Smith, therefore, suggest that the 4-hour threshold lies at or below 87°F (ET).

SIX-HOUR STUDIES

Fine, Cohen and Crist (ref. 10) gave 10 military subjects 6½ hours total exposure to each of four effective temperatures: 65°, 69°, 81°, and 93°F (see line F in figure 10). However, performance testing on the mental task was terminated after 6 hours of exposure, so this will be cited as a 6-hour study. Every week for 4 weeks the same 10 subjects were exposed to the 4 temperatures, 1 temperature per day for 4 successive days. The order of administration of the four temperatures differed with each replication so as to minimize systematic bias from practice, fatigue, and other temporal effects. Either the orders were randomly selected or else followed a Latin square design. Subjects performed an anagram task at the beginning of each session (Trial I) and again 5½ hours later (Trial II). Exactly 35 minutes was allowed for each anagram task. The remainder of each session was spent in performing a discrimination task (not discussed in this review), competing with each other in the verbal game of Ghost, or in resting and eating. The anagrams used were constructed from 1300 of the most frequently occurring three, four, five, six and seven

letter words in the English language. The letters in each word were scrambled. Fine et al report that lists were originally constructed so as to contain 35 words, but that scoring of the first week's data revealed several subjects had completed all the anagrams on the lists. Subsequently, 24 lists were composed, each containing 42 anagrams. The authors do not specify whether the assignment of the new lists of anagrams to the remaining treatment replications was a random process or not, but we presume that it was. Although the results for the three remaining replications did not represent a complete counterbalancing (if indeed a Latin square was used), they were nevertheless averaged in an effort to minimize any effects due to order of treatments or to differences in list difficulty. These results are given in table 5, which shows the average number of correct anagrams for each of the four levels of temperature. A drop in correct solutions is evident at 93°F (ET) on Trial I, but no drop occurs on Trial II. This is surprising, of course, since the longer the exposure period, the greater the effect an extreme temperature should have on performance. The authors report that an analysis of variance revealed a significant trial by temperatures interaction, thus confirming the differential effect of 93°F on Trials I and II. (The level of significance was not reported.) When a separate analysis of variance was applied to each trial, a significant temperature effect ($p < .05$) was obtained for Trial I but not for Trial II! The authors explain their anomalous result as follows:

"... there was some doubt as to whether the significant conditions effect was due to the conditions or to differences in list difficulty. The lists in Trial I of each of the three replications of the 95°/92°F condition (93°F effective temperature) had means of 31.9, 29.1, and 34.7 correct solutions. The mean of 29.1 was lowest of all means obtained and the mean of 31.9 was among the lowest. (The other means ranged from 31.1 to 38.1 with a grand mean of 34.7.)

"As mentioned above, scores were averaged over the replications to minimize the effect of chance variations in list difficulty. However, it is possible that by chance the 95°/92°F condition (93°F effective temperature) was assigned two of the more difficult lists. No other condition had more than one list with a mean under 33.0 (ref. 10, p. 176)."

Thus, the authors lean toward an interpretation in favor of the null hypothesis. The data, however, do not justify an interpretation either for or against the null hypothesis. The facts which

TABLE 5
*Average Number of Correct Anagram Solutions Under
Each Temperature Condition*

Cited from Fine, et al (ref. 10, p. 176)

Trial	Temperature			
	65°	69°	81°	93°
I	34.3	34.7	34.2	31.9
II	35.4	34.8	34.2	35.3

Fine et al marshal to support the null hypothesis can be used just as effectively to argue for the alternative hypothesis that temperature does affect performance. The fact that two of the lowest means occurred under the highest temperature condition on Trial I is exactly the information needed to confirm the alternative hypothesis! Furthermore, it can just as easily be argued that

if there were any differences in list difficulty they were probably differences which obscured the effects of the 93°F temperature on Trial II! The fact is that failure to equate lists prior to the experiment (or to control for their differences in either the design or statistical analysis) makes it absolutely impossible to draw any conclusions on the basis of this study alone. One cannot adequately speculate as to whether an artifactual decrement occurred at 93°F on Trial I or whether a real decrement occurred on Trial II but was washed out by differences in list difficulty. We can say, of course, that all the evidence from other studies would suggest that a real decrement had occurred on Trial II but had been masked by differences in list difficulty. We have seen that Viteles and Smith (ref. 25) found a significant decrement at 87°F for a 4-hour exposure, Mackworth (ref. 17) and Pepler (ref. 20) found decrements at 87.5°F for a three-hour exposure, and Carpenter (ref. 5) found a decrement at 92°F even after a one-hour exposure!*

But on the basis of the Fine et al, study alone, no reasonable conclusion may be drawn because of the failure to equate the anagram lists. Thus, rather than concluding with the authors that there was no decrement in this study, we prefer to conclude that a satisfactory test of the hypothesis was not made. The data from this study, therefore, have not been used in estimating an upper thermal limit for a 6-hour exposure.

EIGHT-HOUR STUDIES

The only eight-hour study is a field study rather than a laboratory study, and it assessed the long-term or cumulative effects of daily, 8-hour exposures. As such, it is not comparable to the other experiments we have reviewed, and so it is not shown in figure 10. Nevertheless we shall review the experiment here, which is by Mayo (ref. 18). He devised an experiment in which two matched groups of U.S. Navy trainees were given classroom instruction in electronics each under different temperature conditions. One group received instruction in a nonair-conditioned building and the other received instruction in an air-conditioned building. Median effective temperatures in the afternoon were 71.3° and 82.0°F, respectively. Corresponding quartile deviations in effective temperature were 2.0 and 1.9. Mayo reports that the median effective temperature was about 2 degrees lower in the morning than in the afternoon in the nonair-conditioned building, but that there was little difference between morning and afternoon temperatures in the air-conditioned building. This suggests that the median daily effective temperatures were actually about 71.3°F and about 81°F.

The two groups of trainees were matched on a variable (unspecified by Mayo) that correlated .62 and .64 with the two measures used in evaluating the effects of classroom instruction. Instructors were matched on the basis of teaching experience and were assigned in such a way as to equalize the level of instruction given the two trainee groups who were undergoing the temperature conditions. All trainees were given 40 hours of instruction per week, so that presumably their duration of exposure was 8 hours per day. After 2 weeks of instruction, the first unit of the course was completed, and an achievement test was administered to both groups. (Each group at this point was composed of 404 trainees.) The second unit was completed 2 weeks later and a second achievement test administered to both groups. However, two classes of 82 trainees (one from each temperature group) were unable to finish the second 2 weeks of training under their respective temperature conditions, so that this left 322 trainees in each group at the end of unit two. The mean achievement test scores on both units for both groups are shown in table 6. On both tests the average score is lower for the group which was given

*The failure to obtain a clear decrement under 93°F after a 6-hour exposure might perhaps be due to the easiness of the task. Under all temperature conditions subjects were apparently unscrambling the anagrams at the approximate rate of one per minute. A more difficult set of anagrams might have provided a better test of impairment in mental performance.

instruction (and presumably testing) in the nonair-conditioned room. When the critical ratio test is applied, neither difference reaches the .05 level of significance, although the difference between the groups on the second achievement test does approach significance ($p < .08$). We must conclude that for 8 hours of exposure a statistically-reliable decrement does not appear at temperatures as low as 81°F.

TABLE 6
*Average Grades Made By Matched Groups Instructed
 Under Two Different Temperature Conditions*

Cited from Mayo (ref. 18, p. 245)

71.3°F			81.0°F			p
Test	Mean	SD	Mean	SD	CR	
1	75.07	9.14	74.76	8.47	0.64	.52
2	72.91	9.25	71.88	10.06	1.76	.08

SECTION III

Conclusion

DETERMINATION OF THE TEMPERATURE-DURATION FUNCTION

Fifteen experiments representing nine different durations of exposure have been reviewed. Only a very small number of studies were represented at any given exposure duration. Nevertheless, at each exposure duration a selection was made of those studies which yielded the most clear-cut results; and then, from those studies the lowest temperature at which a statistically-significant decrement occurred was chosen. This was considered the best-estimate for the performance threshold at that duration of exposure. Each of these values is reproduced in table 7. That is, table 7 recapitulates the conclusions reached in each section of the review of the literature. For each exposure duration, table 7 lists the lowest test temperature at which a reliable decrement was obtained, the study in which the decrement was obtained, the task used in that study, and the level of significance for the difference between that temperature and the control temperature. Close inspection of the table shows an inverse, exponential relationship between exposure duration and lowest temperature yielding significant impairment. This is shown more clearly in figure

TABLE 7
Lowest Test Temperatures Yielding Reliable Decrements

Exposure Duration (mins)	Effective Temperature °F	Experimental Study	Task Affected	Level of Significance
6.5*	114.0***	Blockley & Lyman	Mental Addition and Number Checking	$p < .05$
18.5*	109.0***	Blockley & Lyman	Mental Addition and Number Checking	$p < .05†$
46*	100.5***	Blockley & Lyman	Mental Addition and Number Checking	$p < .01$
60	95.0°	Wing & Touchstone	Memory for Words	$p < .01$
120	89.0°‡	Carpenter	Problem-Solving	$p < .01$
180	87.5°	Mackworth	Telegraphy and Coding Tasks	$p < .05$
240	87.0°	Viteles & Smith	Mental Multiplication and Number Checking	$p < .05$
360	Data not adequate for reaching any conclusion			
480	Data not comparable to those at other exposure durations			

*Estimated duration at which sig. impairment first occurred.

**Estimated effective temperatures as reported by the authors.

†Significant only after one subject's "atypical" data were dropped from the analysis.

‡Interpolated data point.

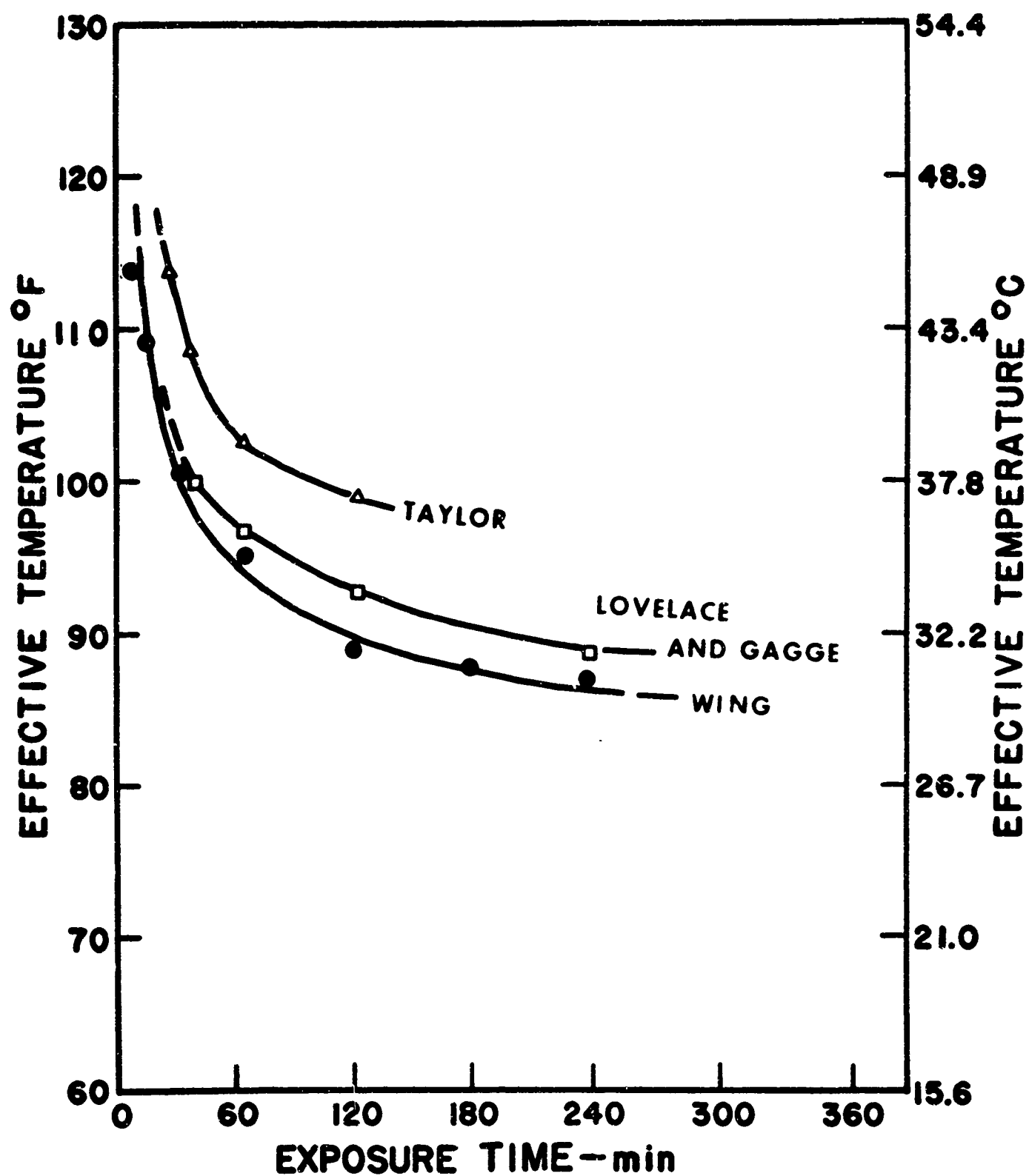


FIGURE 9

Comparison of the Tentative, Upper Limit for Unimpaired, Mental Performance with Both the Tolerable and Marginal Physiological Limits

9 where these lowest temperatures are plotted as solid circles. An exponential curve has been visually fitted to these test points to suggest more clearly the probable shape of the function. It also facilitates comparison of this tentative upper limit for unimpaired mental performance to the upper, physiological, tolerance limits. Two such limits are shown. The first represents recommended or tolerable limits as determined by Lovelace and Gagge. These data are reported in a chart in Connell (ref. 9, p. 79) in terms of dry-bulb, wet-bulb and percent relative humidity readings, but they have been translated here into their effective temperature equivalents. The second limit represents marginal conditions in which collapse is imminent. This marginal limit was determined by Taylor and is also reproduced in Connell (ref. 9, p. 83) in terms of dry-bulb and vapor pressure readings. Again these have been translated here into their effective temperature equivalents.

Comparison of these curves suggests a number of important conclusions. First, the upper limit for unimpaired performance lies below the recommended tolerance limit, and it lies considerably below the marginal or maximum tolerance limit. Comparisons between curves may be made either in terms of temperature or in terms of duration. For example, reading the figure in terms of temperature, we may say that after a 2-hour duration a decrement in mental performance will not occur until about 89°F (ET), a marked physiological impairment will not begin until about 93°F (ET), and imminent collapse will not occur until about 99°F (ET). Reading the figure in terms of exposure time, we may say that at a temperature of 93°F (ET) a decrement in mental performance will not occur until shortly after 1 hour of exposure, a marked physiological impairment will not begin until after 2 hours of exposure, and complete collapse may not be imminent until after some unspecified exposure time (no data available).

SPECIFICATION AND GENERALIZATION OF THE TEMPERATURE-DURATION FUNCTION

There is much potential usefulness to the exponential performance function shown in figure 9, but this should not obscure the fact that it is tentative at best. Ideally, several families of such curves should be presented, each family of curves representing limits for a given set of mental tasks performed by subjects who had undergone specified degrees of training and temperature acclimatization. Not being able to present such data, it becomes especially important to assess what set of conditions the curve in figure 9 most adequately represents and to urge the reader not to generalize to conditions which it does not represent.

First of all, all the studies upon which the points are based (except for the two performed by Mackworth) measured performance during learning. The curve in figure 9 therefore shows the effects of temperature during acquisition of either a new task or of reacquisition of a task not recently, systematically rehearsed. The probable curve for highly-practiced subjects would lie above the present curve; i.e., closer to the tolerable physiological limit.*

Secondly, the studies represent different degrees and types of acclimatization. Most notably, the studies of 2-hours and 3-hours duration used subjects either naturally-acclimatized over a period of six months or else artificially-acclimatized in daily sessions over a minimum period of several weeks. The studies of one-hour duration and less only involved acclimatization accrued during the main experiment which was conducted over a period of just 1 week. Finally, in the Viteles and Smith study of 4-hours duration, only four days of preliminary exposure were held in addition to the main experiment. The effect of these differences in acclimatization on the curve in figure 9 is probably one of depressing the 1-hour and 4-hour thresholds. Thus, for fully

*This has already been demonstrated by Mackworth (refs. 15 and 16) who showed that highly-skilled telegraphers did not suffer impairment at as low a temperature as did telegraphers of average ability.

acclimatized subjects, the actual upper limits for unimpaired performance for 1-hour and 4-hours duration might be slightly higher.

Third, there may be shifts in the curve of figure 9 depending upon the type of mental task employed. As it happens, five of the seven experiments listed in table 7 which showed significant decrements were ones which used mental arithmetic or number-checking tasks. The curve in figure 9 should reasonably represent the upper thermal limit for unimpaired performance on these tasks. As this review has shown, in a number of instances experiments using other tasks appeared to show decrements at roughly the same temperatures. However, information on this point is insufficient. The mental arithmetic task may be particularly sensitive to stress (see Grether, ref. 12). This, then, would suggest that the impairment threshold for other tasks might lie somewhat closer to the physiological tolerance limit, i.e., lie between the present curve and the recommended physiological tolerance limit of Lovelace and Gagge.

Finally, the problem of subject populations should be examined. The experiments listed in table 7 used predominantly military subjects with the exception of the study by Wing. In his study students were used who covered approximately the same age range as military subjects in the other studies described here and had been screened with a flight physical. Nevertheless, differences in subject populations may exist, and it would be desirable to have additional data from a military population on which to base the 1-hour threshold. However, with the exception of the 1-hour threshold, the points on which the curve in figure 9 are based are specific to samples drawn from British and American military populations. These military populations, of course, include various Armed Forces from various nations (U.S. Naval Reserve Pilots, British Naval Ratings, and British Naval Pilots.)

We may summarize by saying that the tentative, upper limit for unimpaired mental performance should not be generalized to all stages of practice, to all degrees of temperature acclimatization, to all types of tasks, or to all subject populations. The curve most adequately characterizes the performance of artificially-acclimatized military subjects on a highly stress-sensitive task either during their learning or else during their re-acquisition of skill on the task.

The effects of either (1) increased training on the task; (2) increased acclimatization to high temperature; or (3) selecting a less stress-sensitive task should be to raise the present curve. This suggests that the present curve describes the lowest temperatures at which decrements will probably appear. If this is the case, the band or area between the tentative upper performance limit and the physiological tolerance limit of Lovelace and Gagge may be viewed as an "impairment zone." Most of the thresholds for various mental tasks would either coincide with the present limit (based primarily on mental arithmetic) or lie above it somewhere in the impairment zone. Most of the thresholds for subjects more highly practiced (and/or more skilled) would lie within the impairment zone and above the present upper limit. Again, the thresholds for subjects who were more completely acclimatized (than the subjects of the studies reviewed here) would also lie within the impairment zone somewhat above the proposed upper limit. This concept of an impairment zone, properly utilized, should help to reduce the tendency to over-generalize the curve shown in figure 9.

Appendix

This appendix contains a re-analysis of experimental data obtained by Givoni and Rim and reported in full by them in their recent article (ref. 11, pp. 104-107). Table 8 shows their data recast into the five class-intervals constructed by the present author (see table 4, page 16). The data entries in this table are the number of mental multiplication problems done by each subject in each run under each of the seventeen test temperatures. The average number of problems done by each subject for all the temperatures represented in a given class-interval are also shown in table 8. These values appear in the lines labelled "Average." The reliability of these averages differs, of course, because the number of test temperatures represented in any given class-interval varies all the way from one to six. Nevertheless, if it is assumed that the reliabilities are such that the rank order of the averages would not change even if more test temperatures were represented in the lowest class-intervals, then the non-parametric, Friedman, two-way analysis of variance may be applied to the ranks of these averages (see Siegel, ref. 24, pp.

TABLE 8
Re-Organization of Givoni and Rim Data on Number of Completed Problems

Class-Interval Midpoints	Test Temperature	Run 1 Subject No.				Run 2 Subject No.			
		A	B	C	D	A	B	C	D
70°F	70.2	16	15	15	14	16	13	16	13
Average =	70.2	16.0	15.0	15.0	14.0	16.0	13.0	16.0	13.0
	73.5	20	15	—	13	19	15	—	13
75°F	74.5	19	17	16	14	18	17	17	15
Average =	74.0	19.5	16.0	16.0	13.5	18.5	16.0	17.0	14.0
	77.9	17	14	15	15	15	14	16	16
	79.0	17	14	—	13	17	13	—	14
80°F	80.0	17	14	13	14	18	15	15	13
	80.9	14	15	13	12	15	13	13	13
Average =	79.5	16.2	14.2	13.7	13.5	16.3	13.8	14.7	14.0
	84.9	19	16	15	14	18	16	16	14
85°F	85.2	20	13	14	14	19	14	16	14
	85.9	11	13	10	12	14	12	9	14
	86.4	17	12	12	12	17	12	13	13
Average =	85.6	16.8	13.5	12.8	13.0	17.0	13.5	13.5	13.8
	88.0	17	15	12	12	20	16	14	12
	88.5	16	13	17	14	16	14	18	17
90°F	89.0	11	9	—	15	12	10	—	17
	89.5	19	15	7	7	20	16	6	8
	89.8	11	12	8	10	12	11	8	10
	90.1	21	16	—	15	22	15	—	15
Average =	89.2	15.8	13.3	11.0	12.2	17.0	13.7	11.5	13.2

166-172). This test determines whether matched subjects perform differently under several (k) experimental conditions as revealed by high or low ranks falling more often than chance under given experimental conditions. In applying this test to Givoni and Rim's data, the situation of matched subjects assigned to k conditions is replaced by the situation of the same subject taking all k conditions. This means that the test as applied here requires the additional assumption that any effects due to sequential testing have been effectively counterbalanced. This assumption, of course, is already implicit in Givoni and Rim's design.

For each subject separately, the averages of the number of problems done under each temperature are ranked. For Run 1 the rankings for all four subjects are as shown in table 9, and for Run 2 they are as shown in table 10. Beneath the columns in both tables are the sums of the column ranks (R_j). If temperature had no effect on problems done, these sums would differ only by chance. The Friedman test consists of computing a statistic, χ^2_r , which is based on the column sums. For a given number of rows and columns this statistic may or may not be significant. For Run 1, $\chi^2_r=301.6$. With five columns and four rows, this value of χ^2_r is significant at $p<.001$. For Run 2, $\chi^2_r=297.3$; this is also significant at $p<.001$. We may conclude that the number of problems completed under the different temperature conditions on both Run 1 and Run 2 significantly differed from chance.

To determine which temperature conditions were responsible for the obtained significance, the nonparametric Walsh test (see Siegel, ref. 24, pp. 83-87) was employed. The Walsh test assumes a symmetrical distribution, that is, a distribution in which the mean and median coincide. The nature of the data obtained by Givoni and Rim suggest that this assumption is adequately met. There is no evidence that subjects were working so close to a performance limit as to skew performance scores. Givoni and Rim's learning curves (ref. 11, p. 113) indicate that there was continuous improvement in performance over the entire seventeen days of testing and that, at best, asymptotic performance was just being approached on the seventeenth day.

TABLE 9
*Each Subject's Average Performance Scores Ranked Across
the Five Temperature Levels for Run 1*

Temperature Midpoints	70°F	75°F	80°F	85°F	90°F
Subject					
A	4	1	5	2	5
B	2	1	3	4	5
C	2	1	3	4	5
D	1	2.5	2.5	4	5
R_j	9.0	5.5	11.5	14.0	20.0

To apply the Walsh test, difference scores (d_i 's) are obtained between the averages of problems done under any two of the temperature conditions. For example, in table 8 there are four d_i 's between the 70° and 75° condition of Run 1, one d_i for each subject: -3.5, -1.0, -1.0, +0.5. The Walsh test determines whether the average of any such set of d_i 's departs from zero by chance or not (at some specified level of significance). With only four d_i 's, all of them must

TABLE 10

*Each Subject's Average Performance Scores Ranked Across
the Five Temperature Levels for Run 2*

Temperature Midpoints	70°F	75°F	80°F	85°F	90°F
Subject					
A	5	1	4	2	3
B	5	1	2	4	3
C	2	1	3	4	5
D	5	1.5	1.5	3	4
R _j	17.0	4.5	10.5	13.0	15.0

lie above (or below) zero to reject the null hypothesis at the .062 level of significance (using a one-tailed test). (This is the highest level at which the null hypothesis may be rejected using a one-tailed test based on four d,'s. Since this is below the previously set level of .05, it will be more appropriate to use the term "almost significant" in referring to this level.) Since in our example only three of the four d,'s have the same sign, we would conclude that the difference in number of problems done under the 70° and 75° temperatures was not significantly different than could be attributed to chance. (A quick way to determine whether performance under any two temperatures differed almost significantly is simply to inspect tables 9 and 10: if all subjects have their highest rank under a given temperature condition, then performance under that temperature differs almost significantly from performance under all the other temperatures (since all four d,'s must then be of the same sign.) This situation is present in table 9, where the 90° temperature contains all ranks of 5. This means that for Run 1 performance was almost significantly poorer under this temperature than under any of the others. For Run 2 (table 10), this was not true. However, it should be noted that the ranks under the 90° temperature are all greater than the ranks under the optimal 75° temperature. Therefore, it can be concluded that on Run 2 performance differed "almost significantly" between the optimal and highest temperature. Inspection will also show that performance under the 70° and 85° temperatures also differed almost significantly from performance under the optimal 75° temperature. We may conclude that the results of these tests tend quite strongly to confirm the results of Carpenter who found by a method of interpolation between test temperatures that a significant decrement in performance occurs between 89-90° F (ET).

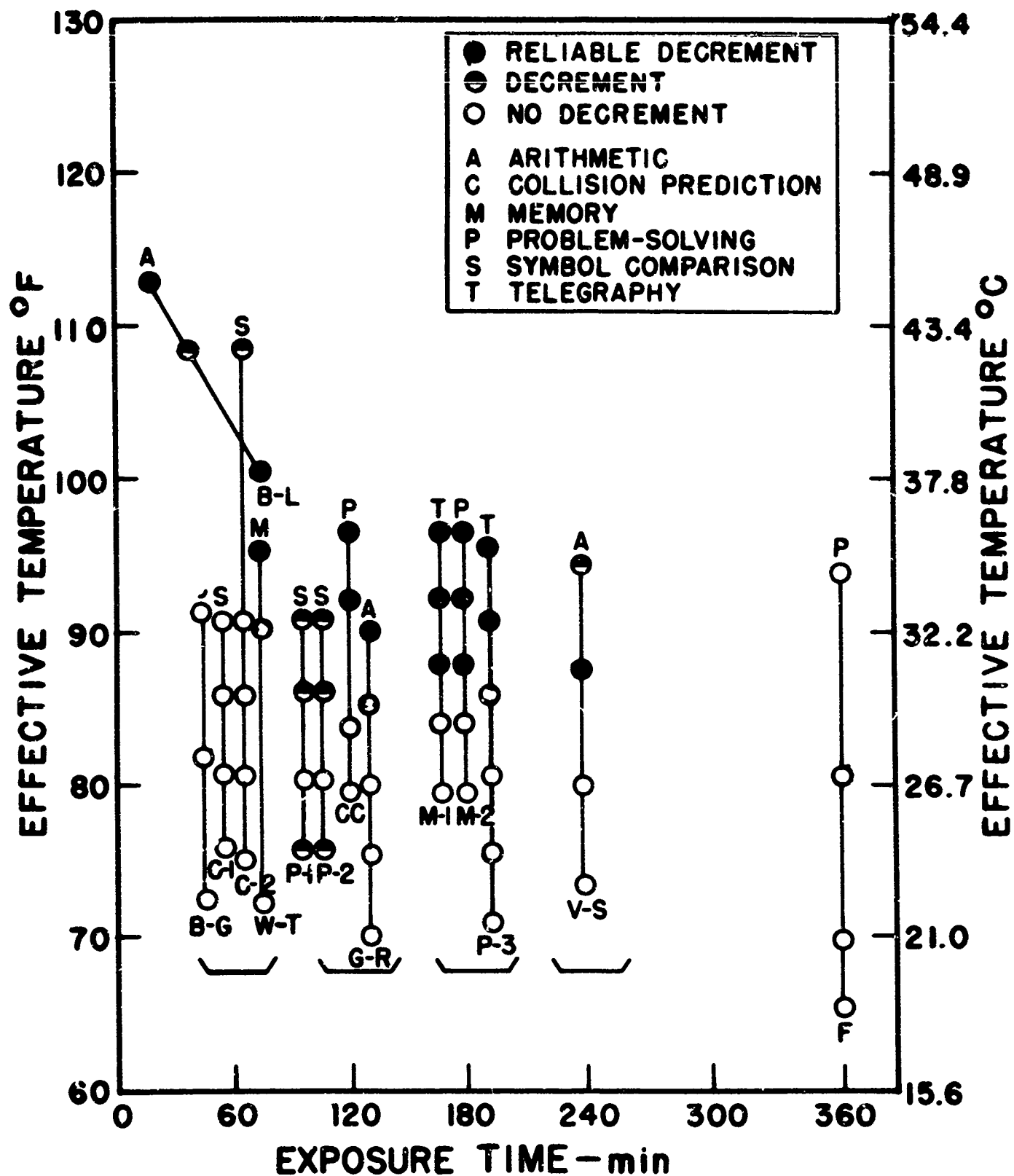


FIGURE 10

Summary of the Exposure Durations and Effective Temperatures
Used in the Fifteen Experiments Reviewed

References

1. Bartlett, D. J., and D. G. C. Gronow, *The Effects of Heat Stress on Mental Performance*, FRPC Report 846, Royal Air Force Institute of Aviation Medicine, England, August 1953. AD 30 748.
2. Bedford, T., *Environmental Warmth and Its Measurement*, War Memorandum No. 17, Royal Naval Personnel Research Committee of the Medical Research Council, London, England, 1946.
3. Bell, C. R., and K. A. Provins, "Effects of High Temperature Environmental Conditions on Human Performance," *J. Occup. Med.*, Vol. 4, pp. 202-211, 1962.
4. Blockley, W. V., and J. H. Lyman, *Studies of Human Tolerance for Extreme Heat. III. Mental Performance Under Heat Stress as Indicated by Addition and Number Checking Test*, AF Technical Report 6022, Air Materiel Command, WPAFB, Dayton, Ohio, October 1950. ATI 95 445.
5. Carpenter, A., *The Effect of Room Temperature on the Performance of the Resistance Box Test: A Performance Test of Intelligence*, Royal Naval Personnel Report 46/318, London, England, August 1946.
6. Chiles, W. D., *Effects of Elevated Temperatures on Performance of a Complex Mental Task*, Wright Air Development Center Technical Report 57-726, WPAFB, Ohio, December 1957. AD 142 192.
7. Chiles, W. D., *Effects of High Temperatures on Performance of a Complex Mental Task*, Wright Air Development Center Technical Report 58-323, WPAFB, Ohio, July 1958. AD 155 811.
8. Chiles, W. D., "Effects of Elevated Temperatures on Performance of a Complex Mental Task," *Ergonomics*, Vol. 2, pp. 89-96, 1958. (Summarizes results of refs. 6 and 7.)
9. Connell, L., *The Effect of Heat Upon the Performance of Men in High-Speed Aircraft: A Critical Review*, SDC Report 151-1-17, Office of Naval Research, Washington, D. C., June 1948. ATI 44 173.
10. Fine, B. J., A. Cohen, and B. Crist, "Effect of Exposure to High Humidity at High and Moderate Ambient Temperatures on Anagram Solution and Auditory Discrimination," *Psychol. Rept.*, Vol. 7, pp. 171-181, 1960. AD 251 918.
11. Givoni, B., and Y. Rim, "Effect of the Thermal Environment and Psychological Factors Upon Subject's Responses and Performance of Mental Work," *Ergonomics*, Vol. 5, pp. 99-114, 1962.
12. Grether, W. F., *Criteria of Anoxia Tolerance: I. Development of Psychological Tests for Use in the Altitude Chamber*, Publ. Board Report No. 15512, Department of Commerce, Washington, D. C., 1942.
13. Houghten, F. C., and C. P. Yaglou, "Determining Lines of Equal Comfort," *Trans. Amer. Soc. Heat. Vent. Engr.*, Vol. 29, pp. 163-176, 1923.
14. Lyman, J. H., *The Performance of Men in Extreme Heat and Environments*, Unpublished Doctor's Dissertation, University of California, 1951.

15. Mackworth, N. H., *Effects of Heat and High Humidity on Wireless Telegraphy Operators Hearing and Recording Morse Code Messages*, Royal Naval Personnel Report 45/243, London, England, October 1945.
16. Mackworth, N. H., "Effects of Heat on Wireless Operators Hearing and Recording Morse Code Messages," *Brit. J. Industr. Med.*, Vol. 3, pp. 143-158, 1946.
17. Mackworth, N. H., Researches on the Measurement of Human Performance," In Sinaiki, H. W., (Ed.), *Selected Papers on Human Factors in the Design and Use of Control Systems*, Paper 7, pp. 174-331, Dover Publications, Inc., New York, N. Y., 1961. (In part summarizes refs. 15 and 16.)
18. Mayo, G. D., "Effect of Temperature Upon Technical Training," *J. Appl. Psychol.*, Vol. 39, pp. 244-246, 1955.
19. Morton, G. M., and J. P. Dennis, *The Effect of Environmental Heat on Performance, Measured Under Laboratory Conditions*, War Office Clothing and Stores Experimental Establishment Report No. 99, London, England, 1960.
20. Pepler, R. D., *The Effect of Climatic Factors on the Performance of Skilled Tasks by Young European Men Living in the Tropics: A Task of Morse Code Reception*, Royal Naval Personnel Report 53/769, London, England, February 1953. AD 22 166.
21. Pepler, R. D., *The Effect of Climatic Factors on the Performance of Skilled Tasks by Young European Men Living in the Tropics: A Complex Mental Task with Varying Speed Stress*, Royal Naval Personnel Report 54/794, London, England, November 1953.
22. Pepler, R. D. *The Effect of Climatic Factors on the Performance of Skilled Tasks by Young European Men Living in the Tropics: A Complex Mental Task with Varying Speed Stress at Two Levels of Incentive*, Royal Naval Personnel Report 54/796, London, England, December 1953.
23. Pepler, R. D., "Warmth and Performance: An Investigation in the Tropics," *Ergonomics*, Vol. 2, pp. 63-88, 1958. (In part summarizes refs. 21 and 22.)
24. Siegel, S., *Nonparametric Statistics for the Behavioral Sciences*, New York, McGraw-Hill, 1956.
25. Viteles, M. S., and K. R. Smith, "An Experimental Investigation of the Effect of Change in Atmospheric Conditions and Noise Upon Performance," *Heating, Piping and Air Conditioning*, Amer. Soc. Heating and Ventilating Engineers, New York, New York, Vol. 18, pp. 107-112, March 1946.
26. Wing, J. F., and R. M. Touchstone, *A Bibliography of the Effects of Temperature on Human Performance*, Aerospace Medical Research Laboratories AMRL-TDR-63-13, Wright-Patterson AFB, Ohio, February 1963.
27. Wing, J. F., and R. M. Touchstone, *The Effects of High Ambient Temperature on Short-term Memory*, Aerospace Medical Research Laboratories AMRL-TR-65-103, WPAFB, Ohio, September 1965.
28. Yaglou, C. P., and W. E. Miller, "Effective Temperature with Clothing," *Trans. Amer. Soc. Heat. Vent. Engrs.*, Vol. 31, pp. 89-99, 1925.

Security Classification

DOCUMENT CONTROL DATA - R&D		
(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)		
1 ORIGINATING ACTIVITY (Corporate author) Aerospace Medical Research Laboratories, Aerospace Medical Division, Air Force Systems Command, Wright-Patterson Air Force Base, Ohio		2a REPORT SECURITY CLASSIFICATION UNCLASSIFIED
		2b GROUP N/A
3 REPORT TITLE A REVIEW OF THE EFFECTS OF HIGH AMBIENT TEMPERATURE ON MENTAL PERFORMANCE		
4 DESCRIPTIVE NOTES (Type of report and inclusive dates) Final report, early 1964 - January 1965		
5 AUTHOR(S) (Last name, first name, initial) Wing, John F.		
6 REPORT DATE September 1965	7a TOTAL NO OF PAGES 32	7b NO OF REFS 28
8a CONTRACT OR GRANT NO b PROJECT NO 1710 c Task No. 171002 d	9a ORIGINATOR'S REPORT NUMBER(S) AMRL-TR-65-102	
9b OTHER REPORT NO(S) (Any other numbers that may be assigned this report)		
10 AVAILABILITY/LIMITATION NOTICES Qualified requesters may obtain copies of this report from DDC. Available, for sale to the public, from the Clearinghouse for Federal Scientific and Technical Information, CFSTI (formerly OTS), Sills Bldg, Springfield, Virginia 22151.		
11 SUPPLEMENTARY NOTES	12 SPONSORING MILITARY ACTIVITY Aerospace Medical Research Laboratories, Aerospace Medical Division, Air Force Systems Command, Wright-Patterson AFB, Ohio	
13 ABSTRACT Fifteen experiments done in various laboratories have assessed the effects of high thermal stress on mental performance. These experiments represent different combinations of exposure time and effective temperature. These studies were reviewed, and the upper thermal limit for unimpaired mental performance was found to vary systematically with exposure duration. Specifically, the lowest test temperatures yielding statistically-reliable decrements in mental performance decline exponentially as exposure durations are increased up to 4 hours. When this temperature-duration curve for mental performance is compared with physio- logical tolerance curves, it is found to lie well below them at every point in time.		

DD FORM 1473

1 JAN 64
AF-WP-8-AUG 64 400

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Stress, psychology Stress, physiology Thermal limits Exposure duration Performance Tolerance Analysis of variance Thresholds, physiology Reaction, psychology Memory Mental tasks						

INSTRUCTIONS

1. **ORIGINATING ACTIVITY:** Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization (*corporate author*) issuing the report.

2a. **REPORT SECURITY CLASSIFICATION:** Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2b. **GROUP:** Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. **REPORT TITLE:** Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. **DESCRIPTIVE NOTES:** If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. **AUTHOR(S):** Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. **REPORT DATE:** Enter the date of the report as day, month, year, or month, year. If more than one date appears on the report, use date of publication.

7a. **TOTAL NUMBER OF PAGES:** The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7b. **NUMBER OF REFERENCES:** Enter the total number of references cited in the report.

8a. **CONTRACT OR GRANT NUMBER:** If appropriate, enter the applicable number of the contract or grant under which the report was written.

8b, 8c, & 8d. **PROJECT NUMBER:** Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9a. **ORIGINATOR'S REPORT NUMBER(S):** Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9b. **OTHER REPORT NUMBER(S):** If the report has been assigned any other report numbers (*either by the originator or by the sponsor*), also enter this number(s).

10. **AVAILABILITY/LIMITATION NOTICES:** Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

- "Qualified requesters may obtain copies of this report from DDC."
- "Foreign announcement and dissemination of this report by DDC is not authorized."
- "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through _____."
- "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through _____."
- "All distribution of this report is controlled. Qualified DDC users shall request through _____."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. **SUPPLEMENTARY NOTES:** Use for additional explanatory notes.

12. **SPONSORING MILITARY ACTIVITY:** Enter the name of the departmental project office or laboratory sponsoring (*paying for*) the research and development. Include address.

13. **ABSTRACT:** Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as (TS), (S), (C), or (U).

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. **KEY WORDS:** Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, rules, and weights is optional.